

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 23, 2003, 07:12:33 ; Search time 4178 Seconds  
(without alignments)  
3963.498 Million cell updates/sec

Title: US-09-805-337A-2  
Perfect score: 3183  
Sequence: 1 MLLFSEVILISWSTVGEG.....SSPPRAICQCKFEYPICE 569

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+P2n.model -DEV=xih  
-O/cgpn2.1/USPTO/US09805337/rnat.17072003.135605.18919/app.query.fasta.1.711  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09805337.@CGN.1.1.2586.@rnat.17072003.135605.18919 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl.\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_in:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pln:\*  
35: em\_hgt\_rdt:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtg\_hum:\*  
40: em\_hgtg\_mus:\*  
41: em\_hgtg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3183	100.0	2823	6	AX274975	AX274975 Sequence
2	3183	100.0	2823	9	AF295327	AF295327 Homo sapi
3	2733	85.9	1707	6	AX274977	AX274977 Sequence
4	1536	48.3	3926	9	HSB	Y00716 Human mRNA
5	1537.5	48.3	3934	4	SSC278470	AJ278470 Sus scrofa
6	1524	47.9	2955	10	MUSCFHRD	M29009 Mouse compl
7	1463.5	46.0	4256	10	RN0320522	AJ320522 Rattus no
8	1455.5	45.7	4229	6	A91883	A91883 Sequence 1
9	1426	44.8	2729	10	AF436847	AF436847 Rattus no
10	1419.5	44.6	4300	10	MUSH	M12660 Mouse Cfh 1
11	1307	41.1	2715	6	A91885	A91885 Sequence 3
12	1284	40.3	2132	9	HUMCHF	M17517 Human compl
13	1054.5	33.1	1264	9	HUMHAB	M65283 Human facto
14	1054.5	33.1	1266	6	BC016735	BC016735 Homo sapi
15	1050.5	33.0	1266	6	AX409739	AX409739 Sequence
16	1050.5	33.0	1266	6	HUMHAAA	M65280 Human facto
17	1038.5	32.6	1148	9	HSB361	X56209 Human mRNA
18	1038.5	32.6	1148	9	HSCOMH362	X56210 H. sapiens m
19	927.5	29.1	1040	6	AX411119	AX411119 Sequence
20	927.5	29.1	1040	6	HSDESK59	X64877 H. sapiens m
21	927.5	29.1	1075	9	BC022283	BC022283 Homo sapi
22	906.5	28.5	1734	10	MUSCFHRB	M29008 Mouse compl
23	861	27.0	2008	4	BTCOMP	X98697 B. taurus mr
24	783	24.6	2358	10	MUSCARB	D10071 Mouse mRNA
25	782	24.6	2377	10	BC030166	BC030166 Mus muscu
26	764.5	24.0	2171	9	HUMFRIITC	M14057 Human facto
27	763.5	22.3	1269	6	AX411132	AX411132 Sequence
28	710	22.3	1269	9	HSDESK59	X68679 H. sapiens
29	710	22.3	1269	9	AL353809	AL353809 Human DNA
30	683	21.5	155892	9	AX147475	AX147475 Sequence
31	681.5	21.4	1929	6	AX411193	AX411193 Sequence
32	681	21.4	1315	6	HSCHFRP4	X98337 H. sapiens m
33	681	21.4	1315	6	BC026782	BC026782 Mus muscu
34	663.5	20.8	3071	10	MUSCFHRC	M29010 Mouse compl
35	631	19.8	2356	6	A91886	A91886 Sequence 4
36	561	17.6	1532	6	BC029390	BC029390 Homo sapi
37	560.5	17.6	1115	9	AL139418	AL139418 Human DNA
38	522.5	16.4	83166	9	HUMHAC	M65294 Human facto
39	488.5	15.3	3377	5	PBXCOMPREG	I21703 Paralabrax
40	476	15.0	186760	2	AC026056	AC026056 Homo sapi
41	460.5	14.5	150626	9	HSJ177P10	AL049744 Human DNA
42	449	14.1	180155	2	AL161735	AL161735 Homo sapi
43	449	13.7	73320	9	HSJ177P10	AL049741 Human DNA
44	437	13.7	73320	4	AF038131	AF038131 Ovis arie
45	435.5	13.7	4020	4	AF038131	AF038131 Ovis arie

## ALIGNMENTS

RESULT 1



Db 1774 GGGAAATTGAATCTATATGTGA 1800

RESULT 2  
AF295327  
LOCUS Homo sapiens complement factor H-related protein 5 mRNA, complete  
DEFINITION AF295327  
ACCESSION AF295327  
VERSION AF295327.1 GI:13195238  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS McRae, J.L., Cowan, P.J., Power, D.A., Mitchell, K.L., Kemp, B.E.,  
Morgan, B.P. and Murphy, B.F.  
TITLE Human factor H-related protein 5 (FHR-5). A new  
complement-associated protein  
JOURNAL J. Biol. Chem. 276 (9), 6747-6754 (2001)  
MEDLINE 21125890  
PUBMED 11058592  
2 (bases 1 to 2823)  
REFERENCE  
AUTHORS McRae, J.L., Murphy, B.F. and Cowan, P.J.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-2000) Nephrology, St. Vincent's Hospital, 41  
Victoria Parade, Fitzroy, Victoria 3065, Australia  
LOCATION/Qualifiers  
1. 2823  
FEATURES  
source  
CDS  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
94..1803  
/note="FHR-5"  
/codon\_start=1  
/product="complement factor H-related protein 5"  
/protein\_id="AAK15619.1"  
/db\_xref="GI:13195239"  
/translation="MLLFVSLISWSTVGEGTLCDFPKIHGFLYDEEDYNPEFQ  
VPTGEVYSCENYVSPKSFMTRTCTEKGMSPTKCLRMCSFEFFVKNHSESGI  
IHLEGTVDIICNTGSLONNEKINISCVRGWSTPTICSTGKECHPILFANVDOP  
KREKYGVDVLKESCKNLIRGSDVOCYFGMSNPFTCGOVSCGPPQLSNGE  
VKEIRKEEYGHNEVEYDGNPFIINGPKICQYDGEEMTLPLCYOVATCCYIPLE  
VGYVSPVPPYGHGVSEYENCNEYAMISNNITCNGITELPKCVATRHOKRCKIA  
GVNITKLKLSGERHNHNSIRIRCSDIYRHRISVCINKMNEVYCTEKRRCFPP  
POIPNAQNNMTTVTVNODKEKVAVLCKENYLEPAKIVCKDKRMOSLPKCVSTACV  
PPSINNGTTSFPLSVPPGSTVTVRCQSFYKLOGSVYTCGRNKMSPSPCLDVCV  
VSEENMKNKNIQIKWRNDGKLYAKTDGDAVEFOCKPRHKAMISPPRALCOBEKEFEP  
ICE"

polya\_signal 2705..2710  
BASE COUNT 960 a 531 c 546 g 786 t  
ORIGIN

Alignment Scores:  
Pired. No.: 5,49e-276 Length: 2823  
Score: 3183.00 Matches: 563  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-805-337a-2 (1-569) x AF295327 (1-2823)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
DB ATGTGCTCTTATTCAGTGAATCCTAATCTCATGTGGTATCCACTGTGGGGAGAGGA 153

QY 21 ThrLeuCysAspPheProIysIleHisGlyPheLeuTyraSpGluIuaSpIlyraSn 40  
DB ACACCTTGATGATTTCCAAAATAACACATGATTTCTGTATGATGAAGAGATTATAC 213

QY 41 ProPheSerGlnValProThrGlyGlyIuaValPheTyTrSerCysGlyIuaSpPheVal 60  
DB CTTTTCCTCCAACTCTACAGGGGAGATTCTTATTAATCTCTGTGATTAATTTTGTG 273

QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluIlyTrpSerProThr 80  
DB TCTCTTCAAAATTCCTTTGGACTCCGCAAAACATGCACAGAGAGAGATGGTCACCAACA 333

QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100  
DB CCGAAGCTCTCAGATGTGTTCTCTTCTTGTGAAAAAGTCATCTCATCTCAATCTCA 393

QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyIysSerLeu 120  
DB GGACTAATACATCTGGAGGTGATGATGACAAATTTTGGACACAGATACACCTT 453

QY 121 GlnAsnAsnGlyLysAsnIleSerCysValGluArgIlyTrpSerThrProIleCys 140  
DB CAACAAATAGAGAAACAACTTTCGTGTAGAACGGGGCTGGTCCCTCTCCCATATCC 513

QY 141 SerPheThrLysGlyGlyCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160  
DB AGCTTCCTAAAGGAGAAATGTCATGTTCCAAATTTTGAAGCCCAATGTAGATGCTCAGCCA 573

QY 161 LysLysGluSerThrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180  
DB AAAAAAGAAAGCTACAAAGTTGAGACGCTTCAAAATTCCTCCGAGAAAAATCTTAA 633

QY 181 ArgValGlySerAspSerValGlnCysTyrglnPheGlyTrpSerProAsnPheProThr 200  
DB AGAGTTGGATCAGACCTCATCTTCAATTTGGGTTGGTGGTCCACTTATCCACACA 693

QY 201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyIuaValLys 220  
DB TGCAAAGAGCAAGTACATCATGTGCTCCACCTCCCAACTGCTCCAAATGTGTAAG 753

QY 221 GluIleArgLysGluIuIuTrpGlyHisAsnGluValGluIuTrpAspCysAsnProAsn 240  
DB GAGATTAAGAAAGAGAAATATGACAAATGAAGTGTGAAATATATGCAATCTCAAT 813

QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro 260  
DB TTATATATTAACGGCGCTTAAGAAATACATGTGTGATGAGAAATGCAACTTACCC 873

QY 261 ThrCysValGluGlnValLysThrCysGlyTyrlleProGluLeuGluIuTyrglyVal 280  
DB ACTTGCTTGAACAAATGGAACATGTGATATATTCAGTCAATCCAGTACGGTTATGTT 933

QY 281 GlnProSerValProProTyrglnHisGlyValSerValGluValAsnCysArgAsnGlu 300  
DB CAGCCGTCTGTCTCTCTCTTATCAACATGCAAGTTTCAGTGCAGGTGAATTCACAAATGAA 993

QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320  
DB TATGCAATGATGGAATTAACATGATTACTGTATTAATGAATATGAGACAGCTTCCCT 1053

QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340  
DB ATGTGTGTTGCAACACACCACTTAAGAGGTGCAAAATAGCAGAGATTATATAAAAACA 1113

QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyraGlyCysSer 360  
DB TTACTCAAGCTATCTGGAGAAAGATTTAATCATTAATCTTCAATACGTTTACAGATCTTCA 1173

QY 361 AspIlePheArgTyraTrpHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380  
DB GACATCTTCAGATACAGGACGCTAGCTGTATTAACGGGAAATGGAATCCTGAAATAGAC 1233

QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400  
DB TGCACAGAAAAAAGGAGAACAAATCTCTCCACCCGACCTCAGATACCTATGCTCAGAA 1293

QY 401 MetThrThrValAsnTyrglnAspGlyGluLysValAlaValLeuCysLysGluAsn 420  
DB ATACACACCACTGATTAATTCAGATGAGGAAAAAGTACGCTTCTCTGTAAAGAAAAAC 1353

OY		421	TyrLeuLeuPProGluAlaValSGluIleValCysLysAspGlyArgTrpGlnSerLeuPro	440
Db		1334	TATCTACTCCAGAAACAAAAGAATTGTATGTAAAGTTCGCACATGCCAATCATTTACCA	1413
OY		441	ArgcysValGlnSerThrAlaTyrCysGlyProProPheSerIleAsnAsnGlyAspThr	460
Db		1414	CGCGTGTTGGAGTCTACAGCATATTGTGGCCCCCTCCATCTGATTAAACAATGGAGATACC	1472
OY		461	TherSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer	480
Db		1474	ACCCATATCCCATTTATCCAGTATATCTCCAGGGTTCACAGTAGACCTACCGTTGCCAGTCC	1533
OY		481	PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGlnPro	500
Db		1534	TTCTATAAACCTCCAGGGCTCTGTAAAGCTTAACATGCACAAATAAACAGTGGTCANACCA	1593
OY		501	ProArgCysLeuAspProcysValValSerGluGlnAsnMetAsnLysAsnAsnIleGln	520
Db		1594	CCAAGATGCCAGATATCCATGTGTGTGAATCTAGASAAACATGAACAAAAAATACCTACAG	1653
OY		521	LeuArgSTPAArgAsnAspGlyLysLeuArgAlaLysThrGlyAspAlaValGlnPheGln	540
Db		1654	TTAAATATGAGAAACAGATGAGAAACTCTATTCGCAAAAACAGGGGATGCTGTGAATTCAG	1713
OY		541	CysLysPheProHISLysAlaMecIleSerSerProProPheArgAlaIleCysGlnGlu	560
Db		1714	TGTAAATTCACCATTAAGCATGATATATCATCACACCATTTCCAGACATCTGTCAGGAA	1773
OY		561	GlyLysPheGluTyrTyroIleCysGlu	569
Db		1774	GGGAATTTGAAATATCTATATGTGCAA	1800
RESULT 3				
AX274977		1707 bp	DNA	linear
LOCUS				
DEFINITION		Sequence 3 from Patent WO0168695.		
ACCESSION		AX274977		
VERSION		AX274977.1		GI:16547599
KEYWORDS				
SOURCE		unidentified.		
ORGANISM		unclassified.		
REFERENCE		1		
AUTHORS		Murphy, B.F.		
TITLE		A novel factor-h related protein 5 and antibodies thereto		
JOURNAL		Patent: WO 0168695-A 3 20-SEP-2001;		
FEATURES		Baxter International Inc. (US)		
source		Location/Qualifiers		
		1..1707		
		/organism="unidentified"		
		/db_xref="taxon:32644"		
	misc_feature	1..1707		
		/note="Generic sequence"		
BASE COUNT		340 a 170 c 280 g		686 others
ORIGIN				
Alignment Scores:				
Pred. No.:		7.48e-236		Length: 1707
Score:		2733.00		Matches: 475
Percent Similarity:		83.48%		Conservative: 0
Best Local Similarity:		83.48%		Mismatch: 94
Query Match:		85.86%		Indels: 0
DB:		6		Gaps: 0
US-09-805-337A-2 (1-569) x AX274977 (1-1707)				
OY		1	MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGluGly	20
Db		1	ATGYTNTNTNNTTYWSNGTNATHYTNAHTMSNTGGGTNMAGCNGTGNGMGARGCN	60
OY		21	ThrLeucyAspPheProLysIleHisIsticLysPheLeuTyrAspGluGluAspTyrAsn	40
Db		61	ACANTNTNGAYATYYCCAAARATACAYAGGNTTYYTNTTAYGARGARGAYTAVAY	120

Oy	41	ProPheSerGluValProThrGluValuAlaPheIleuTYrSerCysGluTYrAsnPheVal	60
Db	121	CCNTTWNCARGTNCNCACMGNGKARGTNTTYATYAWSTGTGCATATAAATYTGCTN	180
Oy	61	SerProSerGlySerPheTrpThrArgIleThrCysThrGluGluGlyTYrPseProThr	80
Db	181	WSNCCNMNAARWSTNTYTGACNMGNNTHTACNTGYACGARGAGCGTNGGWSNCCNACN	240
Oy	81	ProGlyCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer	100
Db	241	CCNNAATCYTNNMCNATGTGWSNTTYCCNTTYGTNMAARAAGCCAAWNSCARNSMNSN	300
Oy	101	GlyLeuIleIleIleGluGlyAspThrValGlnIleIleCysAsnThrGlyTYrSerLeu	120
Db	301	GGNTTNAHCAAYTTGARGNGGATVACNTNCARATHAHTGTGAAATACGNGNTAAYMSTNTN	360
Oy	121	GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProlIecys	140
Db	361	CARAAAYAYGARARAARAAATHTMSNTGYGTNGARFMKNGGNTGTGWSNACCNCNCNATHTG	420
Oy	141	SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro	160
Db	421	WSNTTYACNNAARGNGARTGYCAIGTNCNCNATHYNGRCNAAAYGTNGATGCCNARCNCN	480
Oy	161	LysLysGluSerTYrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle	180
Db	481	AARARGARWRWNTTYAARGTNGCNGNAYCTNTNNAARTYWSNTGMAAARAAYTNAHT	540
Oy	181	ArgValGlySerAspSerValGlnCysTYrGlnPheGlyTrpSerProAsnPheProThr	200
Db	541	MGNGTNGCNSNGATWSNGTNCARTGYATACRTTYGGNTGWSNCCNAAATYTYCCNACN	600
Oy	201	CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysValLys	220
Db	601	TGYARAGNCARGTNGMWSNTGYGCGNCCNCCNCAARTYWSNAAAYGANGCARGTNAAR	660
Oy	221	GluIleArgLysGluGluIuTYrGlyHisAsnGluValValGluTYrAspCysAsnProAsn	240
Db	661	GARATHMGNARGARGARTAGCNCVAAAYCARGTNGTNGARTAAAGATGYAAAYCCNAA	720
Oy	241	PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro	260
Db	721	TTYATHATHAAAYGGCNCNAAARAARHTHARGTGYGTNGAYGGNGARTGGACNACNTTNCN	780
Oy	261	ThrCysValGluGlnValLysThrCysGlyTYrIleProGluLeuGluTYrGlyTYrVal	280
Db	781	ACNTGYGTNGARCARGTNAARACNTGYCGTNAAYATHCCNGARTYTGCAATVAGCNTA	840
Oy	281	GlnProSerValProProTYrGlnHisGlyValSerValGluValAsnCysArgAsnGlu	300
Db	841	CARCCNWSNGTNCNCCNTAYCARATGCGNTGMSNGTNGARGTNAATVGTGMAAYGAR	900
Oy	301	TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro	320
Db	901	TAYCNCNTGATHGCGNAAATATATGATHCNTGYATHAAAYGGCATHTGGACNCAARTNCN	960
Oy	321	MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr	340
Db	961	ATGTGCGTNGCNCACAYCARYTNAARBNGTGYARARHTGCCNGCTNAAAYATHAARCN	1020
Oy	341	LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTYrArgCysSer	360
Db	1021	YTNTNNAARTYNNWSNGNNAARGARTTYATCAAYAAWNSMGNATHIMGTATIMGNTGYMNSN	1080
Oy	361	AspIlePheArgTYrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp	380
Db	1081	GAYATHHTYMGNTTYMCGCAWMSGNTNGYATHAAAYCGNARBTGGAAYCCNCACTNCAY	1140
Oy	381	CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn	400
Db	1141	TGYACNARARARMKNGARCAATYTYTGCCNCCNCCNCCNARATHCCNAAAYGCCNARAY	1200

OY		401	MethrhhrrhThvLaSnTYGhaSpGLVgluIysValAlaValLeuCYalysGluAsn	420
Dd		1201	ATGCACNACMACNGTNAAYTATCAGAAAGGNARARAGTNGGNTYNTTGTAAGAARAY	1260
OY		421	TyIdeLeuPProGluAlaLysGluIleValCysLySaSpgLYatGTpGlnSerLeuPro	440
Dd		1261	TAYYTNTNCNGARGCNAAARGAATHGTNTGYAARGAYGGNMGTGCCARMSVTTCNN	1320
OY		441	ArgysValJlGuserrThrAlaTyrcysGlyProProPserLTleasnaSngLYaspThr	460
Dd		1321	MGRNYGTNGARMSNACGCTATGTGGCCNCCNCCMSMAHAAYAAAGGGAACN	1380
OY		461	ThSerPheProLeuSerValTy-ProProGlySerThrValThrTyArgcysGlnSer	480
Dd		1381	ACMNSNTTYCCNTYNMSGTTATCCNCCNCGNMSNMCNGTNACTAYMGWTGTCARSN	1440
OY		481	PheTyTrLysLeuGInGlySerValThrValThrCysArghaSnlysGlnITrpSerCLeuPro	500
Dd		1441	TTYTAYARARYTNCARGMWSNGTINACNTGNMGMAAYARCAATGGMSGARCCN	1500
OY		501	ProArgcysLeuAspproCysValValSerGlnGluInsMetLeaSnlysaSnTIleGln	520
Dd		1501	CANNGCNTGYTYNGAVICCTGTGTGTMSSNCRARAPAYAGAAVAAAATAATTCAR	1560
OY		521	LeuLSTTPRARgaSaSpgGLyLSLeuTyralAlaLysThrGlyAsPaLaValAlGlupheGln	540
Dd		1561	YTNAARFGGMGMAAYGAVGGNAARYTMTAVGCNABACNONGNAVGGCTGTGATTTCAR	1620
OY		541	CysLysPheProHISylsyalamellieserSerProPropieatrgalaIIEcysGInGU	560
Dd		1621	TGYARTTYCCNCAVYAARGCATATATHSMNSMNCNCCNTTYMGNGCATHTGYCARGAR	1680
OY		561	GlyLysPheGluTyrrProIIEcysGlu	569
Dd		1681	GGNAARTTYGARTRAYCCNATHHTGYGAR	1707
RESULT 4				
HSH				
LOCUS	HSH	3926 bp	mRNA	linear PRI 08-APR-1997
DEFINITION	Human mRNA for complement factor H.			
ACCESSION	Y00716 M32093 X07525			
VERSION	X00716.1 GI:31964			
KEYWORDS	complement factor H; glycoprotein; regulatory protein.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 3926)			
JOURNAL	Ripoche,J., Day,A.J., Harris,T.J. and Sim,R.B.			
MEDLINE	The complete amino acid sequence of human complement factor H			
REFERENCE	Biochem. J. 249 (2), 593-602 (1988)			
AUTHORS	88134059			
TITLE	2 (bases 1 to 3926)			
JOURNAL	Day,A.J.			
COMMENT	Direct Submission			
	Submitted (23-FEB-1988) Day A.J., MRC Immunohistochemistry Unit, Dept.			
	of Biochemistry, University of Oxford, South Parks Road, Oxford,			
	Ox1 3OU			
	On Jul 25, 2002 this sequence version replaced gi:183745.			
	see Immunogenetics 27:211-214(1998) concerning TYR <-> HIS at AA			
	384			
	see X07523 for truncated (1.8 kb) mRNA sequence.			

FEATURES  
SOURCE

CDS

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="B-38-1, R2a and PE3"
/tissue_type="liver"
/clone_lib="human liver cDNA in pGL3/PvuII/8"
74..3769
/codon_start=1
/product="factor H"
/protein_id="CAA68704.1"

```

```

/db_xref="GI:319668"
/db_xref="SWISS-PROT:P08603"
/translation="MRLAKITCLMLAICVAEDCNELPPRNREITELGSMSDQTYP
GTQALYKRCRPGYRSIGNVIMWCRCRGEWALNLRRCOKRCPCHPGDTPRGTTLNGG
VEFYVCAAVYTCNEGYQLLGEINTRECECDGTNDIPLICEVVKLEPVTPENGKIVS
AMEPREYHFGQAVFYCNCSGKYEKTEGDEMGSDGQPMKKEPKCEISCKSPDVIIN
SPIQKIIYKKEBERFYCRNMGEYKSEGDACATCSGMPRLPSCEKSCSDNYPIMGV
LYPLDKRRGTGDEITYCCRCNGRYPRATRCRGTCTGTGTPARCITLPCDYDIDIKG
LYHEMRRNRPFPYVANGKYYTYTCDEHFETPSSYDHIHCITQDMSAPVCLKCTF
YLENENGNHGRKEVQGSIDVACHPGVALPFAQTYTTCMEGMSPTPCLIVKTES
SSIDLENGFISEQYTVALKERAKYQCKGYVADQGETSGSIRCKDGMASDPTCLIK
GDIPYEMARTEKNDFTWPKLNDITLDECHDGYESTNGSTTGSIVCGYNMSDLPLIC
RECELPKIDHLVPRKDDQYKGVGVLEKSCPKYPTIYQPNVQCYHKLSPDLPLIC
BOYQSCGPPPELNLNGNVEKTEKEEYGHGSEVATYCNPRIMKGPKNICQVDEMTLL
VCIVPESTICGDIPELEHMAOILSPYYGDSVERNCSPTMIGHRTTCHGWPTT
LPQVADIRLKKCKKSNLILDEHLKNNKEPDHNSVIRCRCKEWMHTVCLNGRM
PEVNSMAOILCPRPPOIPNSHNMTTLYNDGKRVSVLCQENTLIDGEEITCKD
RMOSIPLCVERKIPCSOPPOLEHGTINSSHSQSEVSHVGLKSLYTCGGRISIEENTC
CYMGSSPPQCEGLPCKSPPEISGVAHMSDSVOYGEELVYKCEGEGIDGPAIA
CLGEKNSHPSCITKIDCLSLPFEENAIPEGKKDQYKAEQVATYTCATYRKMDGASN
TCINSRMTGPRTPCRDTCNVNPTQNAVIVSPKMSKDYKPGCEVATYTCATYRKMDGASN
VMCLGNMTTERPQCKDSTNGKCCPPPIDNGDITTSPLSYAASVETOCNIOLE
NKRITCRKGQMSERPCKLPCVYSREIMENITAIARMTAKOYLISFTGSEVFCRKR
YRLSSRTPLRTTCWDKLEYPTCAKR"
sig_peptide
128. 1276
74. 1127
/product="factor H"
variation
994
/note="polymorphism"
/replacement="c"
1277
/replace="t"
misc_feature
1409
/note="divergence point of full length (4.3 kb) and
truncated (1.8 kb) mRNA species"
variation
2881
/note="polymorphism"
/replacement="g"
BASE COUNT 1306 a 689 c 855 g 1076 t
ORIGIN
Alignment Scores:
Pred. No.: 5.18e-128 Length: 3926
Score: 1538.00 Matches: 298
Percent Similarity: 43.25% Conservative: 93
Best Local Similarity: 32.96% Mismatches: 155
Query Match: 48.32% Indels: 358
DB: 9 Gaps: 3
US-09-805-337A-2 (1-569) x HSH (1-3926)
QY 23 CysasppheProLysIleHisHisGlyPheLeuTYrAspGluGluAspTYrAsnProPhe 42
|||||:||||| ||| ||||| ||||| |||
1046 TGTGATTATTCACGACATTAAACATGAGAGTCTATATCAACAGAAATGTGGTAGACCATATAC 110
43 SerGlnValProThiGlyGluValPheTYrTYrSerCysGluTYrAsnPheValSerPro 62
||||| ||| :||| :||| ||||| :|||
1106 TTTCACGTAGCTGTGAGAAATATATATCTCTATTACTGTGATGACAACTTTTGGAGCTCGG 116
QY 63 SerLysSerPheTrpThrArgIleThiTYrCysTnngGluGlyTYrPseProThrProLys 82
|||||:||||| ||| |||||:||||| |||
1166 TCAGGAAGTACTGGGATTCACATTCATTGCACACACAGATGATGTCGCCACAGATACCA 122
DB 1226 TGCCCTACGAAATGTATTTCCTTATTTGGAAATATGATATAATCAAAATCATCGAAGA 128
QY 83 CysLeuArgMetCysSerPheProPheValIlySasngIyHisSerGlnSerGlyLeu 102
||||| ||| |||||:||||| |||:||||| |||
1226 TGCCCTACGAAATGTATTTCCTTATTTGGAAATATGATATAATCAAAATCATCGAAGA 128
QY 103 IleHisLeuGluGlyAspPheValGlnIleIleCysAsnThiGlyTYrPseLeuGlnAsn 122
||||| ||| :||| :||| ||||| |||:||||| |||
1286 AAGTTTGTACAGGGAATATCATATACAGCTTGCCCTGACATCTGCTACAGCTCTCTCAAAA 134
DB 123 AsnGluTYrAsnIleSerCysValGluArgGlyTrpSerThiProProIleCys----- 140

```

```

Db      1346 GCGCAGACCACGATTACATGTATGGAGAAWGGCTGTCTACTCTCCAGATGCATCCGT 1405
OY      140 -----
Db      1406 GTCAAAACATGTTCCAAATCAAGTATAGATATGAGATGGGTTATTTCTGAATCTCAG 1465
OY      140 -----
Db      1466 TATACATATGCTTAAAGAAAAAGGAATATCAATGCAAACTAGCATATGTAACAGCA 1525
OY      140 -----
Db      1526 GATGTTGAACATCAGAGATCAATTAGATGTGGAAAGATGATGTCAGCTCAGCCACAG 1585
OY      141 -----
Db      1586 TGCATTAAATCTTGTGATATCCAGATTTATGATGCCAGAACATAAATGACTTCACA 1645
OY      143 -----
Db      1646 TGGTTTAAGCTGAATGACACATTTGACATGATGCATGATGTTATGAAGCAATACT 1705
OY      143 -----
Db      1706 GGAAGCACCACGTGTTCCATAGTGTGTGTTACAAATGTTGTTGATTACCACATATCT 1765
OY      144 ---LysGluGluCysHisValProIleLeuGluAlaAsnValAspIleProLysLys 162
Db      1766 TATGAAGAAGATGCGAACTTCTTAATAATGATGTAACACTTATGTTCTCATGCGCAAGAA 1825
OY      163 GluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182
Db      1826 GACCGATTAATAAGTTGGAGAGGTTGAATTTCTCTCCCAACCGAGATTACAAATAGTT 1885
OY      183 GlySerAspSerValGluCysTyrGlnPheGlyTyrPserProAsnPheProThrCysLys 202
Db      1886 GGACCTAAATTCGTTACGTGCTACCACTTGGATTGTCTCTGACCTCCCAATATGATAA 1945
OY      203 GlyIleValIArgSerCysGlyProProGlnLeuSerAsnGlyIleValLysGluIle 222
Db      1946 GAGCAAGTACATCATGTTGCTGCACCTCTGAACCTCCTCAATGGGAAATTTAAGAAAA 2005
OY      223 ArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsnPheIle 242
Db      2006 ACGAAAGAAATATGACACAGTGAAGTGGTAATATTATTGTCATCTGATTTCTTA 2065
OY      243 IleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrPheThrLeuProThrCys 262
Db      2066 ATGAAGGACCTAATAAATTCATATGTGTGATGAGAGTGCACACTTTACACAGTCT 2125
OY      263 ValGluGlnValLysThrCysGlyTyrIleProGlnLeuGluTyrGlyValGlnPro 282
Db      2126 ATTGGAGAGAGAGACTGCTGGAGATTAACCTGAACCTTGAAACATGGCTGGGCCAGCT 2185
OY      283 SerValProProTyrGlnHisGlyLysValIleSerValGluValAsnGlyArgAsnIleTyrAla 302
Db      2186 TCTTCCCTCTTACTACTAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 2245
OY      303 MetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrPheThrLeuProMetCys 322
Db      2246 ATGATTTGGACACAGATCAATTAAGTATTCATGAGATGAGATGAGATGAGATGAGATG 2305
OY      323 ValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeu 342
Db      2306 GTGGCAATAGATAAATTAAGAAGTGCAAATCATCAAAATTAATTAATTAATTAATTA 2365
OY      343 LysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIle 362
Db      2366 TTAATAAACAAAGAAATTCATATTAATTCATTAATTAATTAATTAATTAATTAATTA 2425
OY      363 PheArgTyrArgHisSerValCysIleAsnGlyLysTyrPheAsnProGlnValAspCysThr 382

```

```

Db      2426 GAAGATGATACACACACTGTGATAAATGAGAGATGGATCCAGAACTGACTCA 2485
OY      383 GluLysArgGluGlnPheCysProProGlnIleProAsnAlaGlnAsnMetThr 402
Db      2486 ATGGCACAATATACAAATTAATGATCCCACTCCACCTCAGATTTCCAAATTTCCACATATGACA 2545
OY      403 ThrThrValAsnTyrGlnAspGlyLysValAlaValLeuCysLysGlnAsnTyrLeu 422
Db      2546 ACCACACTGAATTTATCGGATGAGAGAAAAGATGATCTGTTCTTGGCAAGAAATATGTA 2605
OY      423 LeuProGluAlaLysGluIleValCysLysAspGlyArgTyrPheIleSerLeuProArgCys 442
Db      2606 ATTCAGGAAGAGAAAGAAATTAATGACAAAGATGAGATGGACATATACACTCTGT 2665
OY      443 ValGlu 444
Db      2666 GTTGAATAAATTCATGTTCCACACCACTCAGATTAAGAACGGAACCTTAATTCATCC 2725
OY      444 -----
Db      2726 AGGTCTTCAAGAAAGTTATGACATGGAGTAAATGATTAATCTTGAGGGTGTCT 2785
OY      444 -----
Db      2786 TTTCAGATTTCTGAAGAAATTAAGAACACATGCTACATGGGAATGACATTCTCACCT 2845
OY      444 -----
Db      2846 CAGTGTGAAGGCTTCTGTTAATCTGCACCTGAGATTTCTCATGCTGTGTACTCTGAC 2905
OY      444 -----
Db      2906 ATGTCAGACGTTATCAGATATGAGAGAAAGTTAGTACAAATGTTTGAAGTTTGA 2965
OY      444 -----
Db      2966 ATTTGATGGCTTCGACATTTGCAAAATGCTTAGAGAAAAATGCTTCACCTTCATCATGC 3025
OY      444 -----
Db      3026 ATTAATAACGATTGTCTGCTGATTACCTACCTTTGAATAATGCCATGCCATGGAGAGAG 3085
OY      444 -----
Db      3086 AAGATGTGTATTAAGCGGGTGAGAGAGTACTTACCTTTGTCAACATATTACAATAATG 3145
OY      444 -----
Db      3146 GATGAGCCAGTATATGTAATGATGATTAATAGCAGATGAGACAGAAAGCCACATGCACA 3205
OY      444 -----
Db      3206 GACACCTCCTGTGTGAATCCGCCACAGTACAATAATGTTATATAGTGTGAGACAGATG 3265
OY      444 -----
Db      3266 ACTAATATCATCTGTGTGAGAGAGTATTCAAATGTAAGAGCCCTTATGAAATGTTT 3325
OY      444 -----
Db      3326 GGGGATGAAGAGTATGTTAATGSAATCGAGCGAAGCCACCTCAATGCAAGAT 3385
OY      445 SerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAspThrThrSerPhePro 464
Db      3386 TCTACAGAAATATGGGGCCCCCTCACCCTATTGACAAATGGGAGATTAATTCATTTCCCG 3445
OY      465 LeuSerValTyrProProGlySerThrValIleTyrArgCysGlnSerPheTyrLysLeu 484
Db      3446 TTGTAGATATATGCTCCACCTTCATCAGTGTAGTACCAATGCCAAGCTTGATCAACT 3505
OY      485 GlnGlySerValThrValThrCysArgAsnLysIleTyrPheSerGluProArgCysLeu 504
Db      3506 GAGGATACAGCAGATTAATGATGAGAAATGACATGCTCAGAACCCACCAAAATGCTTA 3565

```

OY		505	AsprocysvalvAlserGIugluasMecAasnlysnksnslneGInleUylstrparg	524
Db		3666	CATCGGTGGTGAATAATCCGCAGAAATTATTAACCATGATTAAGTTGGTGACA	3625
OY		525	AasnspglylsLeauTYrAlatalsThrglYAspaIalValglupheGlnCyslvsPphePro	544
Db		3626	GCCAAGAAGAAAGCTTTATTCGAGAACAGGTGTAATCACTTAGATTGTGTGMAAGGGGA	3685
OY		545	HislYsalameIIeSerSerrProPhearghAlatlEcysGlngJuglyLvsPphcglu	564
Db		3686	TATGCTCTTCATCACACTTCACACATGTCGAGAACACATCTTGATCGGAACACTGGAG	3745
OY		565	TyrProIlecys	568
Db		3746	TATCCAACTGT	3757
RESULT 5				
LOCUS		SSC278470		
DEFINITION		SSC278470	3934 bp mRNA Linear MAM 27-JUN-2002	
ACCSSION		AJ278470	Sus scrofa mRNA for putative porcine complement factor H (PFH-11 gene).	
VERSION		AJ278470.1	GI:21464593	
KEYWORDS		PFH-11 gene; porcine complement factor H.		
SOURCE		pig.		
ORGANISM		Sus scrofa		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
TITLE		1 Hegazy,G., Willhoelt,U., Manjo,S., Zipfel,P.F. and Hellwege,J. Porcine complement regulator factor H: Molecular cloning, functional characterization and relevance to cross-species transplantation		
JOURNAL		Unpublished		
REFERENCES		2 (bases 1 to 3934)		
AUTHORS		Hegazy,G.		
TITLE		Direct Submission		
JOURNAL		Submitted (16-JUN-2000) Hegazy G., Research Group of Biomolecular Medicine, Bernhard Nocht Institute, Bernhard-Nocht-Str. 74, Hamburg D-20359, GERMANY		
FEATURES				
source		Location/Qualifiers		
		1..3934		
		/organism="Sus scrofa"		
		/db_xref="taxon:9823"		
		/clone="pfh-11"		
		/sex="female"		
		/tissue_type="liver"		
		/dev_stage="adult"		
		/tissue_1fb="porcine liver unizap ghu"		
		/country="Germany"		
		1..3705		
		/gene="PFH-11"		
		1..3705		
		/gene="PFH-11"		
		/function="complement regulatory protein"		
		/codon_start=1		
		/product="putative porcine complement factor H"		
		/protein_id="CAC8199.1"		
		/db_xref="GI:21464594"		
		/translation="MRCPAKFVWLILMTVCVAEDCKREPPRKETEIVSGSNTEQTPPEPTQTAVKCRPGYRTLGSLIVWCXDKGNWLSHPSTICRRKPCGHADTSFGSFHLAVGTOEFGKARVVYTCDEGYOMLEIRNREDNTNNIPLCEVVKCLPVTEPNRNVISVGASOEFKGVAYTVOCYPGSYPDMDLLCTENMGSPPKCIIVKTOKASDIEINGENSENIIFYPLNKOTQYCKRGVYADGATGLITLOSMSARVPCIXSKSDALLENRAIKSDGWTFRLINDVDVECOOGFEGRNGDGASTVGEDGMDSDPACYEEECITGPSCSYLDAPKKHKFKYGVLTKHCRRNRMLCADSVQCIFHECMNPULPICAKGOVKSAPPQLLNGEVKEIQGETOHSEVEVYCNPRLMKSHRIQCVDEWTLPVLCIEERIKGDIPIIDLHGVRHSAPYHHNGSVEGCANAFSIYIGRSITLTCSKWTDOL		

```

sig_peptide      1..54
                 /gene="pFH 11"
mat_peptide      55..3702
                 /gene="pFH 11"
                 /product="porcine complement factor H"
gene             join(1..54,55..3702)
                 /gene="pFH 11"
repeat_region    55..240
                 /rpt_family="short consensus repeat (SCR) #1 [or CCP or
                 sushi unit]"
                 /rpt_unit=56..240
                 241..426
                 /rpt_family="SCR #2"
                 /rpt_unit=241..426
                 427..618
                 /rpt_family="SCR #3"
                 /rpt_unit=427..618
                 619..789
                 /rpt_family="SCR #4"
                 /rpt_unit=619..789
                 790..963
                 /rpt_family="SCR #5"
                 /rpt_unit=964..1152
                 /rpt_family="SCR #6"
                 1153..1329
                 /rpt_family="SCR #7"
                 1330..1512
                 /rpt_family="SCR #8"
                 1513..1689
                 /rpt_family="SCR #9"
                 1690..1866
                 /rpt_family="SCR #10"
                 1867..2052
                 /rpt_family="SCR #11"
                 2053..2232
                 /rpt_family="SCR #12"
                 2233..2409
                 /rpt_family="SCR #13"
                 2410..2592
                 /rpt_family="SCR #14"
                 2593..2799
                 /rpt_family="SCR #15"
                 2800..2973
                 /rpt_family="SCR #16"
                 2974..3150
                 /rpt_family="SCR #17"
                 3151..3324
                 /rpt_family="SCR #18"
                 3325..3507
                 /rpt_family="SCR #19"
                 3508..3702
                 /rpt_family="SCR #20"
3' UTR          3706..3924
polyA_signal    3924..3934
BASE COUNT      1254 a      774 G      896 G      1010 T
ORIGIN
Alignment Scores:
Pred. No.:      5,76e-128      Length:      3934
Score:          1537.50        Matches:      300
Percent Similarity: 42.59%      Conservative: 88
Best Local Similarity: 32.93%      Mismatches:  156
Query Match:    48.30%         Indels:      367
DB:             4              Gaps:        4

```

```

US-09-805-337A-2 (1-569) X, SSC278470 (1-3934)

OY 23 CysAspPheProIysIleHisHisGlyPheLeuTyrAspGluGlnAspTyrAsnProPhe 42
Db 973 TGTACCCGTCGAGTCTATAACATGGAAGCTATAT-----TATGATCATGTGTGATAC 1026
OY 43 SerGlnValProThrGlyGlnValPheTyrTyrSerCysGluTyrAsnPheValSerPro 62
Db 1027 TTTCCAGCAAAATGTAGACAGATATTTTACTACTACTACTGATCATATTTTGGACTGCC 1086
OY 63 SerIysSerPheThrPheThrArgIleThrCysThrGluGluGlyTyrSerProThrProIys 82
Db 1087 TCACGCCGTCGCGGAGTACTACTGACGCAAAACGGAATGGATGGCTCTCAGAAAGTCCA 1146
OY 83 CysLeuArgMetCysSerPheProPheValIysAsnGlyHisSerGluSerSerGlyLeu 102
Db 1147 TGCCTCAGACATGTTATTTTCTTATTTTGGAAATGAGACACACACATCATTTCCAGAAA 1206
OY 103 IleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGlnAsn 122
Db 1207 AATATTTTACAGGCTGAACCTGTAAGAGTTCACTGCTATCCCGCTACAGTCTCCAGAT 1266
OY 123 AsnGluIysAsnIleSerCysValGluArgGlyTyrPheThrProProIleCysSerPhe 142
Db 1267 AATCAGGACACGCTGCTGTGTACAGAGAAATGGCTGTCTCTCCCAAGTCATCTTT 1326
OY 143 ThrIys----- 144
Db 1327 GTCAAAACATGTGCAAAATCAGATATAGAAATTTGAAATGGATTTTTTCTGCAAAATATA 1386
OY 144 ----- 144
Db 1387 TTTACATATCCCTTAATTAACAAACACATATTAAGTGAACAGAGATATGTAAACAGCA 1446
OY 144 ----- 144
Db 1447 GATGGCGCACTTCAGAGATTAAATCAGATGTCTGCAAGATGATGGTCACTGCAGCTGTG 1506
OY 144 ----- 144
Db 1507 TGCATTAACTCTGGAGTGCAGCACTACTGAGAAATGCCAAGATCAAGAGTGTATGACAG 1566
OY 144 ----- 144
Db 1567 TCGTTTAACCTCAATCACACAGTGCAGTGTGAATGCCACAGAGGTTCGAAAGCCGAAT 1626
OY 144 ----- 144
Db 1627 CGAGACACTGCAGAGTCCATTGTGTGTGTGTAAGACGATGGTCTGATAGCCTGTGT 1686
OY 145 -----GlyGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnProIysIys 162
Db 1687 TACGAAGAGATGTATTTACTCTCGAATAAAAAGCTATTTAGATGCTCAACCAACAACT 1746
OY 163 GluSerTyrIysValGlyIysValIleuIysPheSerCysArgIysAsnIleuIleArgVal 182
Db 1747 AACAAATTCAAAGTTGGAGGCGTGTGAATTCACCTCCAGAGAAAGAGCTGTGATGCTT 1806
OY 183 GlySerAspSerValGlnCysTyrGlnPheGlyTyrPheProAsnPheProThrCysIys 202
Db 1807 GGAGCGATTCAGTTCAATGCTTACCACTTTGGATGCTCCCTTAATCTTCCAAACATGTAA 1866
OY 203 ---GlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyIleValIysGln 221
Db 1867 GGAGGCAAGTAATATCTGCTCCACCCCTCAACTCTCAATGGGGGAAATTAAGAA 1926
OY 222 IleArgIysGluIleTyrGlyHisAsnGlnValValGluTyrAspCysAsnProAsnPhe 241
Db 1927 ACACGAAAGAGATATCAACACAGTGTGTGCAAFACGTTTGGCAATCTTCAATTT 1986
OY 242 IleIleAsnGlyProIysIysIleGlnCysValAspGlyIleTyrPheThrProThr 261
Db ----- 261

```

---

```

Db 1987 CTGATGAAGGCTCTCATATAAATTCATGTGTGATGAGAAATGACAAACTTGCCTGTG 2046
OY 262 CysValGluGlnValIysThrCysGlyTyrIleProGlnLeuGlnIleTyrGlyValGln 281
Db 2047 TGTATTGAGGAGAGCGTACCTGTGAGATATTCCGACCTCGACCATGCTATGTCGG 2106
OY 282 ProSerValProProIleGlnHisGlyIleSerValGlnValAsnCysArgAsnIleTyr 301
Db 2107 CATTTGCCCTCCCTCCATCACCATGAGACCTCAGTGAAGTTCGGGCTGCAATGCATCAT 2166
OY 302 AlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrPheIleuProMet 321
Db 2167 TCACGTGTGGACCCAGGTCAATCACCTGTATTAGGGAAGTGAAGTCCCACTTCTCTCG 2226
OY 322 CysValAlaThrHisGlnLeuIysArgCysIysIleAlaGlyValAsnIleIysThrLeu 341
Db 2227 TGCCTTGCACAGATTAACCTCAAGAAAGTAAAGCCTCAAAATATTTCATCTGAGGA 2286
OY 342 LeuIysLeuSerGlyIysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAsp 361
Db 2287 AACCTGTGTGATTAAGACCCGAAATTTGACATTAACACCAACCAAAAGTTACAAATGTCGGGA 2346
OY 362 IlePheArgTyrArgHisSerValCysIleAsnGlyIysTyrAsnProGluValAspCys 381
Db 2347 AATCAGAAAGCAAAATCTCACTCTGTATTAATGAGATGAGATCCAGACTTAACCTGC 2406
OY 382 ThrGluIysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMet 401
Db 2407 AAGAAAGAGATTAATGATTCATCCACCTCCACCTCAATTCCTCAATGCTCAAGATAG 2466
OY 402 ThrThrThrValAsnTyrGlnAspGlyIleIysValAlaValLeuCysIysGluAsnTyr 421
Db 2467 ACAACTACAGTGAATATATAGATGAGAGAAAAAATATATATCTCTGTCGCAAAAACTAT 2526
OY 422 LeuLeuProGluAlaIleIysGluIleValIysIysAspGlyArgTyrGlnSerIleuProArg 441
Db 2527 ATATATTCAGATGAGAGAAATTTGTGTCAAGATGAGAGATGGCAGTCAAGTCCACGC 2586
OY 442 CysValGlu----- 444
Db 2587 TGTGTGAATAAATTCATGTTCTGACACACCTGAAATAGACCATGAAACATTCAGTCA 2646
OY 2647 CCTGCTCTGCAGAAAGAAAGAGAGAAATGTTTGAGCCCAAGAAATATGACATGCAACC 2706
OY 444 ----- 444
Db 2707 AGTTAAGTTATACTTGTGAAGATGTTTGCAGATATCTGAAAGAGATATATGCCAC 2766
OY 444 ----- 444
Db 2767 ATGGAAATGAGATTTCTTGGCTCAGTGTGTGGTGTGCTTGTGCACACACCTGAG 2826
OY 444 ----- 444
Db 2827 ATTCAAGACGTGTGTGTGCTCAGAGAAAGACACTACTGTAGCAGAGAAAGATTACC 2886
OY 444 ----- 444
Db 2887 TACAGCTGTGATGAAGTTTGGGATGTATGACCTGCATCTTAAGATGTTTAGAGAA 2946
OY 444 ----- 444
Db 2947 AATGCTCCCATCTCTCAGAAATGATTAACACGAGATTTGTTGACTTACCCACTTCAT 3006
OY 444 ----- 444
Db 3007 GACCTGTAGTATAGGCCAAAGAGCCATCTGATAAGTCAAGAGACAGTGACTTTT 3066
OY 444 ----- 444
Db 3067 AATGCTTGCATATTTTCAATTGGAAGAGACCTAACACCATAGGTTGATTTAGAGCAGA 3126

```



```

OY 444 ----- 444
Db 3127 TGGATAGGAAGCCAGCATGCAGAGATGTTTCTGTGATATCCACCCAGTGTGGAAC 3186
OY 444 ----- 444
Db 3187 GCTGTACTAGATAGAGGTCAAGCTATCAACTGGTGTGAGAGACAGCTATGAGTGC 3246
OY 444 ----- 444
Db 3247 ATGCAACTTACGCCCTTATGSGAGTGAAGTGATGTTTAATGGAAGTTGGACC 3306
OY 445 ----- 445
Db 3307 AATCCACTCATGCAAGACGCTAAGGAATATGGGGCCACCCCACTATAGACAT 3366
OY 458 GlyAspThrThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArg 477
Db 3367 GGAGACAGACCTCATTTCCCATTTACAGCTATCCCGGAAACAGCTCGAGTACCA 3426
OY 478 CysGlnSerPheTyrLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyr 497
Db 3427 TCCCACTCTACTATGAACTTGAGGAGACAGCTCATTAATGTAAGATGACAGCTGG 3486
OY 498 SerGluProProArgCysLeuAspProCysValValSerGluGlnAsnMetAsnLysAsn 517
Db 3487 TCAGAACCCACCAATGCTTACGTAGCTGTAGTTCAGAAATGATGAGAAACAT 3546
OY 518 AsnIleGlnLeuLysTyrArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaVal 537
Db 3547 AACATAGACCTAATAATGAGACAGCAATAAACTTATTCAGAAACAGATGATCTATT 3606
OY 538 GluPheGlnCysLysPheProIleLysAlaMetIleSerSerProPheArgAlaIle 557
Db 3607 GAATTCGGTGTAGACAAAGATATTTAGAGAGACCCCACTACATTCATTCGAGCAACC 3666
OY 558 CysGlnGluLysPheGlnTyrProIleCys 568
Db 3667 TGTCAAGCAAGCAAGTGTCTACTACTGTGT 3699

RESULT 6
MUSCFHRD 2955 bp mRNA linear ROD 12-JUN-1993
LOCUS Mouse complement factor H-related protein mRNA, complete cds, clone
DEFINITION 9C4.
ACCESSION M29009.1 GI:192563
VERSION M29009.1 GI:192563
KEYWORDS Complement factor.
SOURCE Mouse (strain C57/B10.WR) liver, clone 9C4.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2955)
Vik.D.P., Munoz-Canoves,P., Kozono,H., Martin,L.G., Tack,B.F. and
Chaplin,D.D.
Identification and sequence analysis of four complement factor
H-related transcripts in mouse liver
J. Biol. Chem. 265 (6), 3193-3201 (1990)
1689298
COMMENT Draft entry and computer-readable sequence for (J. Biol. Chem.
(1990) in press) kindly submitted
by D.P. Vik, 13-OCT-1989.

FEATURES
Source 1..2955 Location/Qualifiers
CDS /organism="Mus musculus"
/db_xref="taxon:10090"
69..2495
/note="complement factor H-related protein"
/protein_id="AAA37416.1"
/db_xref="GI:309166"

```

```

/translation="MGFCSLLILSLITLAWLSTAKGEVKSCEPFQFKRYLFEEL
RNPVPSIGNKSYKQDNQSPSGFNFNYLCTQVGMRPEVPCVKCEHVHNEEF
AYMRKIVQGGSLAKYQKQGVSLQNGQDNQEMSPKCIKTCQSVDIETEN
GPFSESPRYALNRETSYKQGYVNTNGETSRSLTCLQNGSPQSDCKSEBPPE
NSVTKNNSWPEKRLKIDLECLIGHENETKHKSGSTCTCYVIGMSDPSCEYELCSYPI
LDRLKLVSPKREKRVGDLLEFSCRGSHVNPDSVQCYHEGSPSPKGVASCAQ
PPEIPGDELNGAKRVEYSHGEVYQDYCKRFLKGNKIQCDVMATLTPVCEERT
CGDIPDELHGVSVKSVPEYHGDSDVEFTCAEFETMTGLGSVGLSGKMTQLPKCVARD
OLEKCRVLTSTDIATIKPKRNEFHNSITVYKCRQKQVEHSTCIQNKRNPNCRK
TSCPPRPQIPLNLTIVETTYKYLDGELSTLQDNITLTDQPEMKCRQWQSLPHICG
LPQGPSPSLTRGVLSLEESYOHGEVYTHCSTGFEIDGPATIKEGKWSQSPKIC
TNCQDVLPTLENAIIRGKRRKSYRQGVTPRQSPQPMQSDTVICVNSRMLKPKVC
DSRGKCPPPIDNGDITSLSLPEVPESSVDQCKYLLKGTKITTCRNKMSERP
TCLHACIPEINEMEARKIILKRNRETNISYHSGEDIEFCCKYGHKARSGSPFRKCI
SGTINPQCE"
BASE COUNT 990 a 550 c 577 g 838 t
ORIGIN

Alignment Scores:
Pred. No.: 6, 41e-127 Length: 2955
Score: 1524.00 Matches: 292
Best Local Similarity: 47.02% Conservative: 87
Best Local Similarity: 36.23% Mismatches: 187
Query Match: 47.88% Indels: 240
DB: Gaps: 6

US-09-805-337a-2 (1-569) x MUSCFHRD (1-2955)

OY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTyrValSerThrValGlyGlyGly 20
Db 84 ATGTCCTCTTATCCAAATATCCCTCACTGATGCTTCCACTCTAAAGGGAAGTG 143
OY 21 ThrLeuCysAspPheProLysIleHisIleGlyPheLeuTyrAspGluLysPyrAsn 40
Db 144 AATCTTGTGAATTTTCCAAATATGACGCTGTATTATTAAGAGATCCCTGAGA 203
OY 41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60
Db 204 CCCAAGCTCCCACTATCTATAGAAATAGTACAGCTATTAAGTGACAAACGGGTTTCA 263
OY 61 SerProSerLysSerPheThrPheArgIleThrCysThrGluGluLysTyrSerProThr 80
Db 264 CCACCTTCTGGGCTTTCTGGGACTCTCTGTCGACAGTACAGAGGTGGAGGCTGAA 323
OY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
Db 324 GTCCCATGTCTCGAAGAAATGCTTTTCCATTTATGTCGACATGACAAATTCGACTG 383
OY 101 GlyLeuIleHisLeuGluLysAspThrValGlnIleIleCysAsnThrLysTyrSerLeu 120
Db 384 GAAATAATATATGTCAGAGGTCAGTCTTTAAAGTCCAGTGTATATATGCTATAGCTT 443
OY 121 GlnAsnAsnGluLysAsnLysSerCysValGluArgGlyTyrSerThrProProIleCys 140
Db 444 CAATAATGTCAGACACAAATGATGATACAGAGAATGGTGTGCTCCCTCCCAATGCG 503
OY 140 ----- 140
Db 504 ATCCGTATCAAGACATGTCATGATCATATGAANTTGAAGATGGGTTTTTTCTGAA 563
OY 141 SerPhe----- 142
Db 564 TCTTTTCGTACATATGCTTAATAGAGAAACATCTCATAGATGTAAACAGGATATGTC 623
OY 142 ----- 142
Db 624 ACAAAATAGTGAGAAACGTCAGATCAATTAATGCTTCAAAATGATGATGTCACCTCAA 683
OY 142 ----- 142
Db 684 CCTCATGTATTAAAGCTTGTGAAAGACCAAGATTTTGAGAAATCTGTAACTAAGAAATAT 743
OY 142 ----- 142

```

Db 744 AGTACATGTTTAAAGCTCAACAGCAAAATTAGACTATGATGTCATGTCATGACATGAAAT 803  
 QY 143 -----Thrlsgly----- 145  
 Db 804 GAATATTAACATACCAAGGCTCTATACGCTACTTATTATGATGTCGTGATACGCC 863  
 QY 146 -----GlucyshsvalProileuGlualaasnValaAspaInlpro 160  
 Db 864 TCATGTTATGAATAGATGACAGCGTTTCCCATTTCTAGACGAAACCTGCTGTTCTGCC 923  
 QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuLe 180  
 Db 924 ACAAAAGAAATACAGAGTGTGAGATTGTAGAAATTCCTCCGCTGACAGA---CAC 980  
 QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyrSerProAsnProthr 200  
 Db 981 AGAGTTGGCCAGATTCAGTCAATGTTACACATTGGATGGTCTCTAGTTTCCCTTACA 1040  
 QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysValLys 220  
 Db 1041 TATAAGTGTCAAGTATGATCATGTGCAACAACCTCCGAAATTCCTAATGGGAAATTAT 1100  
 QY 221 GluIleArgLysGluIleTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240  
 Db 1101 GGAGCAAAAGAGTTCAATACAGCCATGGTGGATGATGATGATGCAAAACCTTACA 1160  
 QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValaAsnGlyGluTyrThrLeuPro 260  
 Db 1161 TTCCTACTAAGAGGCCAATAAATCCAGTGTGATGGATGGATGGCAACCTTGGCT 1220  
 QY 261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyTyrVal 280  
 Db 1221 GTATGTTGTGAGAGGAGCAACATGTGGAGACATTCCTGAACATGGCTGCTGTC 1280  
 QY 281 GlnProSerValProProTyrGlnIleHisGlyValSerValGluValaAsnCysArgAsnGlu 300  
 Db 1281 AAGTTTCTGTCCTCCCTACCAACATGAGATTCAGTGGAGTTCCTGCTGTCAGAAACC 1340  
 QY 301 TyrIleMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrPheGluLeuPro 320  
 Db 1341 TTCACATGATGATGACTGGTCACTTCTGCTTACCTTACGAAAGTGGCAACCTGCTGCT 1400  
 QY 321 MetCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValaAsnIleLysThr 340  
 Db 1401 AATGTGTGCAACAGACCACTGGAAGTGTAGAGTCTGAAGTCAACTGACATAGCA 1460  
 QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360  
 Db 1461 GCAATTTAAACCAAAAGAAATCAATTTTCAGCATTAACCTCCACATGATTAACAATCTACA 1520  
 QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValaAsp 380  
 Db 1521 GACAGCAGAGCTATGACACATCAATCTGTATCAATGAAATVGAATCTGTAACCAAC 1580  
 QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400  
 Db 1581 TGTCAAGGAAACATCC-----TGCCCTCTCCACACACAGATTCCAATTCACCTTAGTG 1634  
 QY 401 MetThrThrValAsnTyrGlnAspGlyGluLysValaAlaValLeuLysLysGluAsn 420  
 Db 1635 ATTTGAACACACTGTGAAATCTTGATGTGAGAAAAATVATGCTGTTCTTGCCCAAGCAAT 1694  
 QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrPheLysLeuPro 440  
 Db 1695 TACCTTAACACAGACCCAGAGAAATGATGTCCAAAGATGGAAGTGGCAGTCATTACT 1754  
 QY 441 ArgCysVal----- 443  
 Db 1755 CACTGCATTGAGCTTCTGTGACCTCCACCTCAATTCCTGCTGACTGTTCTCTCT 1814  
 QY 443 ----- 443

Db 1815 GAGCTAGAGAGTTACCAACATGAGGAGAGGTTACATACCATTTGTTTACAGGCTTTGGA 1874  
 QY 443 ----- 443  
 Db 1875 ATTTGATGACACAGCATTTATTAATGTGAAGAGGAAAGTGTCTGACCACCAAAATGC 1934  
 QY 443 ----- 443  
 Db 1935 ATAAAAAGCAATTGTGACGTTTACCACCAATTGAAAAATGCCATATAAGAGAAAGAG 1994  
 QY 443 ----- 443  
 Db 1995 AAAAATCATATAGACAGAGAGAGAACAGTGCATTCAGATGTCATTCCTTATCAAAATG 2054  
 QY 443 ----- 443  
 Db 2055 AATGCTCAGACACGTGACATGTGTAAATACCGGTGATTTGGAAGCCACTATGCAAA 2114  
 QY 444 GluSerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAspThrThrSerPhe 463  
 Db 2115 GACTCAAGAGGAGAAATGTGGCCCTCCTCCACTATTGACAAATGAGACATCCCTTG 2174  
 QY 464 ProLeuSerValTyrProProGlySerThrValIleThrTyrArgCysGlnSerPheTyrLys 483  
 Db 2175 TCATTACAGAAATATGAAACATTCATCATGCTGACTATCACTGACGACAGAAATATATCTC 2234  
 QY 484 LeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrSerGluProProArgCys 503  
 Db 2235 CTTAAGGAGAAACAGCAATATACATGTAGAAATGGATGGTCTGAGCACCACCAATGC 2294  
 QY 504 LeuAspProCysValaValSerGluGluAsnMetAsnLysAsnAsnIleGlnLeuLysThr 523  
 Db 2295 TTACATGTGATGTATATACCAAGAAACATATAGAGAACGACGCAAAATATTTAAATGG 2354  
 QY 524 ArgAsnAspGlyLysLeuThrAlaLysThrGlyAspAlaValaGlnPheGlnLysPhe 543  
 Db 2355 AGACACACAGAAATATTTATTTCCATTGAGGAGGAGATTTGAAATTTGAAATTAAT 2414  
 QY 544 ProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluLysPhe 563  
 Db 2415 GGATATCATTAAGCAAGAGATCACCGCATTTCTGACAAAGTGCATTAAGTGGCACCATC 2474  
 QY 564 GluTyrProIleCysGlu 569  
 Db 2475 AATTATCCCATCTGTGAA 2492  
 RESULT 7  
 RNO320522 4256 bp mRNA linear ROD 04-SRP-2001  
 LOCUS  
 DEFINITION  
 Rattus norvegicus mRNA for complement inhibitory factor H (Fh gene).  
 ACCESSION  
 AJ320522 GI:15485712  
 VERSION  
 Rattus norvegicus complement inhibitory factor H, Fh gene.  
 KEYWORDS  
 Norway rat.  
 SOURCE  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 REFERENCE  
 1 (bases 1 to 4256)  
 Demberg T., Goetze O. and Schiاف G.  
 Rat complement factor H: molecular cloning, sequencing and expression in tissues and isolated cells  
 Unpublished  
 2 (bases 1 to 4256)  
 Demberg T.  
 Direct Submission  
 Submitted (07-AUG-2001) Demberg T., Immunology, Georg August University of Goettingen, Kreuzbergstr 57, 37073 Goettingen, GERMANY  
 FEATURES  
 source  
 1. 4256  
 /organism="Rattus norvegicus"

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="liver"
/clone_lib="lambdab Triplex"
/dev_stage="ages 10 - 12 weeks"
/tissue_lib="Clontech rat liver 5' stretch plus cDNA library"
24..4209
/gene="FH"
24..3374
/cds
/gene="FH"
/codon_start=1
/product="complement inhibitory factor H"
/protein_id="CAC67513.1"
/db_xref="GI:15485713"
/translation="MRLSARILMLIYVCAEDCKGPPRENSILSGSEOLYSE
GTQATYKCRPGRTIGTIVKVCCKNGEWPSPNRCIKRKPCHPDTPFGSRFVAGS
EPEGAKVYTCDEGYQLLEIDYRECDAGWTNDIPICEVVKCLPTELENGRIVSG
AAEPQEVYFGQVREECNSGFKIEGCKEMSCENGSLNENKPCVEISCLPVRVNG
DGIYLPVKENREYQYKQGVYKERDVACTGSGNPPSCSEMTCIPIYINGI
YTPHRIKHRIDEIRYCKNGLYPATRSPVSKCTITGMIPARCSLKPDPEQFHRG
LYEBSRRYPPVPYIGKEYSVYCDNGFPPQSVDYLRCTVNGSREPPRCRQCTFH
YVEGESLMQRRYIEGOSAKYQCSGSLPNGOPTIICTEGMSPPKCVRIKTCVS
SDIEENGFSSESDIYALNRRYRCQGVYNTGELSGIITCLQDQMSPPSCIKS
CDMPYENAMTKNNNTFWKLNKLDYECIHGYENYKHKGSITCTYDQMSPTSCYE
RECSIPILHODLVPPREKVKYVDSLSFSCSGHRVAGADLYCQHFQSPYFPCGE
OYKSCDQPLEIPNGEIKGTKKVEYSGHGVDEYDCKPRFLKPKIKQCYDQKWTLLPI
CYEYERTCGDLDELHSGSVKLSVPPYHHGDSVEPTCTETFTMGHAYVCSIGRTEL
POCVATDLEKCAKASTGIDAIHPNKNENFNFVSYKCRKOYEHSICINGMDP
EONCRNEKRCPPPPQIPNAQVIEITVAKYLDGEVSVLCQGVYLTQGEVAVKCHRG
WSPLRNCEKRIPCSQPKIEHGSIKSPKSEERDLIESSYHGHTFVSVCDDGRIS
BENRTCKMGKMSLPRCVGIPCGPPSPILGIVSHELESYGEVATVNCDEGRID
GPAFLKCVGQWSEPKCIKTPCDNLFPFEIAKPEKKKSVSGSDVAVFCPPPYRM
LMSDILVTCVNTKIGQPYCKNSGVNPPVPAATILTRKTATVSGDKRYKRYCPFE
LFWGSDVPKTKGFWTPEPKCKDSTKCGPPPIDNGDITSLTPYATRSVEYQCO
NYTLKGNKIVTCRNKMSQPTCLHACTIPEDIMEKHITVLRMRNNAKTIYSQSGENI
EFMCKPGYRKFEKSPFPKCIEGHINPTCV"
4194..4199
/polyA_signal
/gene="FH"
4209
/polyA_site
BASE COUNT 1423 a 788 c 892 g 1153 t
ORIGIN
polyA_signal
polyA_site
Alignment Scores:
Pred. No.: 2,84e-121 Length: 4256
Score: 1463.50 Matches: 289
Percent Similarity: 40.908 Conservative: 84
Best Local Similarity: 31.698 Mismatches: 172
Query Match: 45.988 Indels: 367
DB: 10 Gaps: 4
US-09-805-337a-2 (1-569) x RMO320522 (1-4256)
QY 23 CysAspPheProLysIleHisIseGlyPheLeuTyraSPGLuGlnAspTyraSnprophe 42
DB 996 TGTGATTTTCCACAAATTCAGACGTCTGTATTATGAGAAAGCCGAGACCCCTAC 1055
QY 43 SerGlnValProThiGlyLValPheTyrrTySerCysGluTyraSnpheValSerPro 62
DB 1056 TTCCCAAGTACCTATAGGAAGAGATACAGCTATTACTGTGACACACGGGTTTCAACAGCCT 1115
QY 63 SerLysSerPheTrpThrArgIleThrCysThnGluGluGlyTrpSerProThrProLys 82
DB 1116 TCACAGTCATACTGGGACTACCTTCCTGCACACGTAATAGGGTGGAGCCTGAAGTTCCA 1175
QY 83 CysLeuAlaGlyMetCysSerPheProPheValLysAsnGlyHisSerGlnSerSerGlyLeu 102
DB 1176 TGCCTCAGGCAATGATTTTTCATATTGTGGAATATAGGAATCTTTATACGGCAAAAGA 1235
QY 103 IleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrrSerLeuGlnAsn 122

```

```

DB 1236 AGATATATAGAGGCTCAGTCTGCAGAAAGTCACAGTGCACAGTGGCTATGATCTTCCAAAT 1295
QY 123 AsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys----- 140
DB 1296 GGTCAAGATACATATATTATGTACAGAAATGGCTGCTCCCTCCTCCCAATGCTCCCT 1355
QY 140 ----- 140
DB 1356 ATCAAGACTTGTTCAGTATCAGATATAGAAATTTGAAATGGGTTTTTTTCTGAATCTGAT 1415
QY 140 ----- 140
DB 1416 TATACATATGCTCTAATATAGAAAAACCGGTATAGATGTAAACAGGATATGTAACAAT 1475
QY 140 ----- 140
DB 1476 ACCGAGAAATATTCAGAAATTAATTACTTGTCTTCAAGATGATGTGCACCTGCACCTCA 1535
QY 140 ----- 140
DB 1536 TGCATTAACTTGTGATATGCTGCTGATTTGAGAAATGCTATGACATAAGATAATATACACA 1595
QY 140 ----- 140
DB 1596 TCGTTTAACTCAATGACAAATTTAGACTGATGATGATGATGATGATAATAT 1655
QY 141 SerPheThrLysGly----- 145
DB 1656 AATCAATACCAAGGCTCTTAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1715
QY 146 -----GluCysHisValProIleLeuGluLysValAsnValAspAlaGlnProLysLys 162
DB 1716 TATGAAGAAGAAATGACAGATTCCTCCCTGATACCAAGACTTGAATGTTTCCACAGAGA 1775
QY 163 GluSerTyrrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182
DB 1776 GTAAATATACAAAGTTGGAATTCGTTGATTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1832
QY 183 GlySerAspSerValGlnCysTyrrGlnPheGlyTrpSerProAsnProPheProLys 202
DB 1833 GGAGCAGATTTTACTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1892
QY 203 GlyGlnValArgSerCysGlyProProGlnIleuSerAsnGlyLysValGlyGluLeu 222
DB 1893 GCCCAAGTAAATCATGTGACCACTCTGAAATCCCAATGGGAAATGGAATGGAATGGAATGGAATGGAAT 1952
QY 223 ArgGlyGluGluTyrrGlyHisAsnGluValGlnTyrrAspCysAsnProAsnPheIle 242
DB 1953 AAAAAGTTGAATACAGCCATGCTGACGTGGGGAATATGATGCAAACTTACATTTCTA 2012
QY 243 IleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrrThrLeuProThrCys 262
DB 2013 CTCACAGCAACCAATTAATCCAGTGTGTGACGGAAGTGGACAACTTCCCGGATATGC 2072
QY 263 ValGluGlnValLysThrCysGlyTyrrIleProGluLeuGluTyrrGlyValGlnPro 282
DB 2073 GTTGACTATGAGAGACATGTGACAGCTCTCTGACACTTGAGAGATGCGCTGTGCAAGTTA 2132
QY 283 SerValProProTyrrGlnHisGlyValSerValGlnValAsnCysArgAsnGlyTyrrAla 302
DB 2133 TCTGTCCTCCCTTACCATCATGAGATTCAGGAGCTTCACTGTGACAGAAACCTCTGACA 2192
QY 303 MetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuProMetCys 322
DB 2193 ATGATTTGGACATCACTAGTATTTCTCATTTAGTGAAGGTGAGACCAAGCTTCTCATATG 2252
QY 323 ValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeu 342
DB 2253 GTTGCAACAGATCACTGAGAGAAGTAAAGCCGCAAGTCAACTGGCATAGATGCAAT 2312
QY 343 LysLeuSerGlyLysLysLysPheAsnHisAsnSerArgIleArgTyrrArgCysSerAspIle 362
DB 2313 CATCAAAATAGAAATGAAATTTATCTTACTTGTAGTGTAGAGATGACAAAG 2372

```





```

OY 497 TrpSerGluProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLys 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3516 TGGTCTACGCCACCACTGCTTACATGATCTGATACCAAGATATTGTGAAAAA 3575
OY 517 AsnAsnIleGluLeuLysTrpArgAsnAspClyLysLeuValLysThrGlyAspAla 536
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3576 CATATATATAGTTCTCAGATGAGGAAATCAAGATTATTCCTCAATCAGGGGAGAT 3635
OY 537 ValGluPheGlnCysLysPheProHisLysAlaMetLiesSerProPheArgAla 556
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3636 ATTGAATTCATGCTGTAACCTGGATATAGAAATTCAAGAGATCACTCCGTTGTGACA 3695
OY 557 IleCysGlnGluGlyLysPheGluTrpProIleCys 568
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3696 AAGTCATTCAGCGTCACATCAATATATCCCACTTGT 3731

RESULT 9
AF436847 2729 bp. mRNA linear ROD 23-APR-2002
LOCUS Rattus norvegicus complement factor H-related protein mRNA,
DEFINITION complete cds.
ACCESSION AF436847 GI:20271386
VERSION AF436847.2 GI:20271386
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Rat complement factor H-related protein sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Direct submission
JOURNAL Submitted (19-OCT-2001) Nephrology, University of Chicago, 5841 S.
Maryland Ave., Chicago, IL 60637, USA
REFERENCE 3 (bases 1 to 2729)
AUTHORS Ren, G. and Quijg, R.J.
TITLE Direct submission
JOURNAL Submitted (23-APR-2002) Nephrology, University of Chicago, 5841 S.
Maryland Ave., Chicago, IL 60637, USA
REMARK Sequence update by submitter
COMMENT On Apr 23, 2002 this sequence version replaced gi:16551387.
FEATURES
source
location/Qualifiers
1..2729
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_type="glomerular epithelial cells"
/tissue_type="kidney; also found in liver"
50..2149
/codon_start=1
/product="complement factor H-related protein"
/protein_id="AAU25802.2"
/db_xref="gi:20271387"
/translation="MRFSIMLANILITSLTAKEEECVLHVENGYSFLERK
VODSVKNCYKGFMLPMDIITCTESGMTPPICISFTSCSLIDNGFESD
LTVANRRTRYCKGGYGTANGERSGISTGOTGMSAPSCISGDPVFNKSKND
STWRKNDKLDYECHEGTYGENTKHKSGIMKTGMSKPSCEIEGIPILDRLYV
YPRKIKYVGLDKRSKRGARVGPDSVOCDFGSPSPICKGVRSGGQYPELLNG
EIKTKKAEYSHGDVYEDCKRFLKPNKIOCYDGKWTLLPTCVERKRRKCGDLAEI
EHGSKFSPVPHGDVSEFTCAEFYVIGGSSVCSIGMAQLPRCVATQLEKCA
PLAVIATGNKLDKLNHNSISYKCGSGSYGSCINGRMADPEPTCTREKFCPP
PPQIPNAHIVETLYKIDGKSVYICGSGYLTGPEEMVCKGHGMSLPCTAKITCS
QPKRIDHSGITLPRISDAGDAIESSEHEKTTSPSYVDDCFRISKENGVTCHMKMS
SPRCVESTGKCGPPTIENGDTISLPLVAPLSSVEYQCSFYMGLSKRTCKNG
EMEPKCLACVGLTEEMRONTIIPKTKNKIYIOSGDYVEFVCLAYOKAOSPE
FRTOCIDGHINYPICSK"
BASE COUNT 916 a 502 c 561 g 750 t
ORIGIN

```

```

Alignment Scores:
pred. No.: 3 67e-118 length: 2729
Score: 1426.00 Matches: 284
Percent Similarity: 52.57% Conservative: 84
Best. Local Similarity: 40.57% Mismatches: 188
Query Match: 44.80% Indels: 144
DB: 10 Gaps: 11

US-09-805-337a-2 (1-569) x AF436847 (1-2729)
OY 3 LeuLeuPheSerValIleLeuLieserTrpValSerThrValGlyGluGlyThrLeu 22
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 CCTTACGAAATATCTCTCTATCATCATCACTTCTACTGCTTAAGCGGAAGAGAAA--- 127
OY 23 CysAspPheProLysIleHisIleGlyPhe---LeuTrpAspGluGluAspTrpAsnPro 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 128 TGTGTTTGCATATATGGAATGTTATCTTCTTATTCCTGGAAGAAATTTGTACAA 187
OY 42 PheSerGlnValProThrGlyValPheTrpTrpSerCysGluTrpAsnPheValSer 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 188 GATCAGTCTGTCAAGTC-----AACTGTATAAGGCTTCACCTT 229
OY 62 ProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThrPro 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 CCAATATGCTCAG-----GATCAATATATATGTCAGAGAGTGGCGACACCTCAACCC 283
OY 82 LysCysLeu-----ArgMetCysSerPheProPheValLys-----AsnGlyHisSer 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 ATATCATTCCTCTTCGACAGATGCTCAAGAGTCAGATATAGAAATGATATGATTTT 343
OY 98 GluSerSerGlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGly 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 344 TCTGATCTGATTTATACATATATCTTAATAGAAACACGATATGCTGTAACAGCGA 403
OY 118 TySerLeuGlnAsnAsnGluLys-----AsnIleSerCysValGlnArgIlyTrpSer 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 404 TATFACACAGCAAAATGAGAAAGATCGGATCAATCACTTGGCTACAAACTGAGAGTCT 463
OY 136 ThrProIleCys-----GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 464 GCTCAACCTTCTGCTGATTAAGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523
OY 140 ----- 140
Db 524 AATGATACACATCGCTTAACTCAATGCAAAATAGATTATGAAATGTCACATTCGATAT 583
OY 141 -----SerPheThrLysGly----- 145
Db 584 GAAATTAATATATAGCATACCAAGGCTCCATATATGCTACTACTAGATGATGCTGAT 643
OY 146 -----GluCysHisValProIleLeuGluAlaAsnValAspAla 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 644 AAACCTCATGTTATGAATATGAATGAAATGCAAGATTCCTCAATTAAGCCAGCTTATGTT 703
OY 159 GlnProLysLysGluSerTrpLysValAlaLysPheLeuLysPheSerCysArgLysAsn 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 704 TATCCAGAGAAATTAATATATACAGTTGAGATTTGTAATCTCTTGGACGACACAGA 763
OY 179 LeuIleArgValIleSerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPhe 198
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 764 ---CACAGAGTTGACCGCATTCAGTTCAGTAAGCTATGACTTTGATGGTCCCTTACTTTT 820
OY 199 ProThrCysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnIlyGlu 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 821 CGGACATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 880
OY 219 ValLysGluIleArgLysGluGlyTrpLysHisAsnGluValGlnTrpAspCysAsn 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 881 ATAAGGGAAGAAAGAGTGAATACAGCATGATGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 940
OY 239 ProAsnPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThr 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      941 CCTAGATTTCTACTGTAAGGAGGCCCAATTAATCCAGTGTGTGACGGGAAGTGGACACACC 1000
Oy      259 LeuProthCysValaIgluInVallyThrcysGlyTyrIleProgluLeuIgluTyrGly 278
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1001 TTGCCACATGGCTTACAGAGAAAGAAACATGTGGAGACCTCTCAACTTGAAGCATGGC 1060
Oy      279 TyrValGlnProSerValbProProTyrGlnHisGlyValaSerValaIgluValaIscysArg 298
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1061 TGTGCAAGATTCTCTCCCTCCCTACACACAGCATTCAGATGAGATTACTTGTGTGCA 1120
Oy      299 AsnGluTyrAlaMetIleGlyAsnAsnMetIleThrcysIleAsnGlyIleTyrPheIglu 318
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1121 GAACCTTCACAGTGTGTAAGGAGGATGATCTTGTGATGATGAGGAGGAGGAGGAGGAG 1180
Oy      319 LeuProMetCysValaIleThrcysIleLeuLysArgCysLysIleAlaGlyValaIle 338
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1181 CTTCCCTGCATGTCTTCCACACAGTCACTTGAAGAGTGAAGACCGAGACGAGCTGTGA 1240
Oy      339 LysThrLeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArg 358
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1241 ATAGGAGCAAAATTAACCAAGAAAGATGATGATTAATCATATCACTAGCATCACTACAA 1300
Oy      359 CysSerAspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrPheAsnProglu 378
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1301 TGTGAGGAGAGTCAGAGATGATGATCAATCTGATTAATGGAAGATGAGATGCTGTGA 1360
Oy      379 ValAspCysThrGluLysArgGluInPheCysProProProGlnIleProAsnAla 398
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1361 CCACCTGTTCAGAAAGAAAGAAACATCTGTGCTCCCTCCCTCCACAGATTCCAAAATGCC 1420
Oy      399 GlnAsnMetThrThrThrValaAsnTyrGlnAspGlyGlyLysValaIleValLeuCysLys 418
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1421 CACGTATTTGAACCAACACAGTAAATCTGTGATGAGAGAAATATCTGCTCTTCCAA 1480
Oy      419 GluAsnTyrLeuLeuProgluAlaLysGluIleValCysLysAspGlyArgTyrPheIgluInser 438
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1481 GAGGTACCTACTACTGAGGCCCAAGAAATGCTGTGAACATGTAAGTGGACAGTGC 1540
Oy      439 Leu 439
        |||
Db      1541 TTACACAGGCTGCAGGCAAAATTAATCACTGTCCACGCCCTTAATATAGCATGATGC 1600
Oy      439 439
        -----
Db      1601 ATTAAGTACCAGATTTGTCAGATGACAGAGAGATGCAATTGATGACAGACTCATGAA 1660
Oy      439 439
        -----
Db      1661 CACGGACACATTCAGCTATGTCTGTGATGATGCTTCAAGATATCTAAGAAATGGG 1720
Oy      440 440
        -----
        ProArgCysValaGluSerThrAlaTyr 448
Db      1721 GTAACCTGCACATGGGAAATGAGAGCTCTCCCTCGATGTGTGATCAACAGGGA 1780
Oy      449 CysGlyProProProSerIleAsnAsnGlyAspThrThrSerPheProLeuSerValTyr 468
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1781 TGTGGCTCTCTCCCACTATGTAAGAAATGAGACATCACCTCTTCATTAACAGATATAT 1840
Oy      469 ProProGlySerThrValaThrTyrArgCysGlnSerPheTyrLysLeuGlnGlySerVal 488
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1841 GCACCTTAATCTCACTGTAATATGCAATGCAATCTCTATATAATATCGGCGACCAAG 1900
Oy      489 ThrValThrCysArgAsnLysGlnTyrPheSerGluProProArgCysLeuAspProCysVal 508
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1901 AAATATACCTCCGCAAAATGAGAGCTGGAAGAGCCACCAAGCTTTACATGCAATGTGTA 1960
Oy      509 ValSerGluGluAsnMetAsnLysAsnAsnIleGlnLeuLysTyrPheAsnAspGlyLys 528
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1961 TTACAGAAAGAAATTAATGAGACAGACATTAATTCACCAAGAGGAGCAAAAGATTAAG 2020
Oy      529 LeuTyrAlaLysThrValaValaIgluPheGlnCysLysPheProHisLysAlaMet 548
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2021 ATTATATATCAGTCAAGGGATTAATGTTGAATTTGTGTCTTAAGTGCATATCAAAAGGCA 2080

```

```

Oy      549 IleSerSerProProPheArgAlaIleCysGlnGluGlyLysPheGluTyrProIleCys 568
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2081 CAAGCATCACTGAGATTGCTGACACAGTGCATTTGATGTCATCATATATATCCACTGCG 2140

RESULT 10
MUSH
LOCUS
DEFINITION
MUSH
Mouse Cfh locus, complement protein H gene, complete cds, clones
MH(4,8)
4300 bp mRNA linear ROD 22-APR-1994
ACCESSION
M12660.1 GI:193724
VERSION
M12660.1
KEYWORDS
Mus musculus (strain C57/B10.WR) male liver cDNA to mRNA; Mus
musculus (strain C57/B10.WR) male liver DNA; and Mus musculus
(strain C57/B10.WR) male liver cDNA to mRNA.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 4300)
AUTHORS
Kristensen, T. and Pack, B.F.
TITLE
Murine protein H is comprised of 20 repeating units, 61 amino acids
in length.
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 83 (11), 3963-3967 (1986)
MEDLINE
86233353
PUBMED
2940596
COMMENT
Draft entry and clean copy sequence for (1), (unpublished (1986)
by T. Kristensen, 28-JUL-1986.
FEATURES
source
Location/Qualifiers
1..4300
/organism="Mus musculus"
/strain="C57/B10.WR"
/db_xref="taxon:10090"
/sex="male"
/tissue="liver"
<1..3425
/note="H protein"
/number=1
join(101..3425,3474..3853)
CDS
/note="precursor"
/codon_start=1
/product="complement protein H"
/protein_id="AAA37759.1"
/db_xref="GI:387181"
/translation="MRLSARILWLIIMTYCAEDCGPPRENSILLSGWSBOLYPE
GTQATYRCNRYGFTLTIVKVCCKNGWVANSNPSRLCRKPCGHPGDTPEGSFLAVGS
QPEFGAKVYVYDDGYQLGEIDYRECGADWINDIPLDEVKCLPVTLENGRIYSG
AAETDQEVYFGGVYREFCNSGFKIEGHEKIESENLSMSEKRCVEILCTPRVNG
DGLNVKPYKENVRYKCKHGVCYKHEKGDVAVCTGSGMSOPECIEKRCSPYILNGI
LYPHRIIHRSDDEIRYCNVGPYVAVGPDVPSKTPGMIVPCCILPCEDFQKCYRH
LYEESLRNRPYSTGNKYSYKDCNFPSPSGSMYILCTAGCMPEVPCVAKCYRH
IYENGSAIWEKYVQGSLLKVCYNGYSLQNGDMTCENMGWSPSPKILRIKTCSA
SDIHNGFLSESSSYIALNRRTSYRCKGVYNTBEISGISTICLQNGMSPOSCIKS
CDMPVENSITKNTRTWFKLNDKLDYECLVGENEYKHTKGSITCYIYCMSDPSCE
RECSVPLDRKLIVSPRKRKYRGDGLSECHSGHVGPDSPVOCYHFGMSPPGCTKG
OVASCAPLEIILNGEINGAKKYVSGEYVKKDKRPFILKGNKTCQDGNMTLPY
CIEERTCGDIPLEHSGAKSVYPYHEDDSVFICEENFTMGHGSVCSISKATOL
PKCVANDOLEKCVNLKSTGIEAIKPLTEFTHSTADYCRDQOEYERISICNGKMP
EPNCTSTSCPPVLPNTQVETTVYLETWLDGKLSVLQDENVLTQSEEMVACDGHMO
SLPRCIEKIPCSQPPITIEHGISILPSSSEBRSDSISSSEHETTSYACDDGFRIPE
ENRTIYKMGMSPPRCVGLPGCPSPISPLGYSLSEYOHGEETVYSNGFGIDG
PAFIIQEGKMSPPKICITQCDVYPTVKNAILRGKSKSRGTGROVPRCOSPYOMN
GSDTYCVNSRNATGQVYCDKNSCVDPHPHNPADNITRTKNKYLHCDRVYECRKLDEL
FGQEVWCENGIATEKPKCRDSTGKCGPPPDNIGDITSLSLVEPLSSVEYQCKY
YILKGRKITTCTNGKMSSEPTCTHACVIEIDENMESHNILKMHTEKIYSHSGDIEF
GCKRYGKARDSPPFTCKINGTINPTCV"
101..154
/note="H protein signal peptide"
join(155..3425,3474..3850)
/product="complement protein H"
/note="H protein"

```



Intron 3426..3473  
 /note="H cds intron"  
 exon 3474..4300  
 /number=2  
 BASE COUNT 1401 a 814 c 902 g 1183 t  
 554 bp upstream of XhoI site; chromosome 1.  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,59e-117 Length: 4300  
 Score: 1419.50 Matches: 283  
 Percent Similarity: 39.33% Conservative: 82  
 Best Local Similarity: 30.50% Mismatches: 178  
 Query Match: 44.60% Indels: 385  
 DB: 10 Gaps: 5

US-09-805-337a-2 (1-569) x MUSH (1-4300)

QY	23	CysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsnProPhe	42
DB	1073	TTGGAATTTCCACAAATTCAAAATGAGACGCTGTATATGACAGAGCCTGAGACCCAAAC	1132
QY	43	SerGlnValProThrGlyGluValPheTyrSerCysGluTyrAsnProPheValSerPro	62
DB	1133	TTCCCGATATCTTATAGCAAAATAGTACAGCTATAGTGACACGGGTTTTCACCACT	1192
QY	63	SerLysSerPheThrPheArgIleThrCysThrGluGluGlyTyrSerProThrProLys	82
DB	1193	TTGGGATATTCGGGACTACCTTCGTCGACAGCACAAAGGGTGGAGCGCTGAAGTCCCA	1252
QY	83	CysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerGlyLeu	102
DB	1253	TTGGGATAAAGGATGTTTCCATTTATGAGAAATGAGACTCTGCATCTGGAATAA	1312
QY	103	IleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGlnAsn	122
DB	1313	GTTATGTCGAGGCTAGCTCTTAAAGTCCAGCTGTACAAATGCTATAGCTTCAAAAT	1372
QY	123	AsnGluLysAsnIleSerCysValGluArgGlyTyrSerThrProProIleCys	140
DB	1373	GGTCAGACACATGACATGTACAGAAAGTGGTGGTCCCTCCCAAAATGCATCCGT	1432
QY	140	-----	140
DB	1433	ATCAGACATGTTCCACCATCAGATATACATGACAAATGATTTCTTCTGAAATCTCT	1492
QY	140	-----	140
DB	1493	TCTATATATGCTCTAATATAGAGAAACATCCTATAGATGTAAGCAGGATATGTGACAAAT	1552
QY	140	-----	140
DB	1553	ACTGAGAAATATCAGAGTCAATTAAGTCCCTCAAAATGATGATGCACCTCAACCTCA	1612
QY	140	-----	140
DB	1613	TGCATTAACTCTTGATATGCTGCTATTTGAGAAATCTATACTAAGAATACTAGACA	1672
QY	140	-----	140
DB	1673	TGCTTAAGCTCAATGACAAATTAAGCTATGAATGCTGTTGGATTTGAAATGATAT	1732
QY	141	SerPheThrLysGly-----	145
DB	1733	AAACATACCAAGGCTCTATACATGTACTTATATGATGATGCTGATACACCTCATGT	1792
QY	146	-----GluCysHisValProIleLeuGluAlaAsnValAspAlaInProLysLys	162
DB	1793	TATGAAGAGAAAGCGTGTCTCCACTTACAGACCGAAACTAGTCTTTCCCCAGAAAA	1852
QY	163	GluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal	182
DB	1853	GAATAATACAGACTTGAGACTTTGTTGGAATTCCTGCCATTCAAGAA---CACAGAGTT	1909

QY	183	GlySerAspSerValGlnCysTyrGlnPheGlyTyrSerProAsnProPheProThrCysLys	202
DB	1910	GGGCCAGATTCACTGCAATGCATCAACCTTTGGATGCTCCCTGTTTCCCTCATATATA	1969
QY	203	GlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyValLysGluIle	222
DB	1970	GTCAGATGACATCATGTCACCACTCTTGAATTTCTTAATGGGAAATTAATGACACA	2029
QY	223	ArgLysGluGluTyrGlyHisAsnGluValValGluTyrAspCysAsnProAsnProPheIle	242
DB	2030	AAAAAGTTGAATACAGCCATGCTGAGTGAATATGATGCAAACTGATTCCTCA	2089
QY	243	IleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeuProThrCys	262
DB	2090	CTGAAGGAGACCAATTAATTCACGTCGTGTGATGGGAATTTGGACAACTTGCTGTATCT	2149
QY	263	ValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrValGlnPro	282
DB	2150	ATTGAGAGAGAGAGAAACATGTGAGACATTCCTGAACCTGAACTGCTCCCAAGTGT	2209
QY	283	SerValProProTyrGlnHisGlyValSerValGlnValAsnCysArgAsnGluTyrAla	302
DB	2210	TCTGTTCTCCCTACCACTGACATGACATTCAGTGAAGTTCATTTGTCAAGAAACTTCA	2269
QY	303	MetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeuProMetCys	322
DB	2270	ATGATGTGACATGCGTCAGTCTTCTTCATAGCGAATAATGACCACTGCTCTTAATCT	2329
QY	323	ValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeu	342
DB	2330	GTTGCAACAGCCCAACTGAGAAAGTGTAGAGTCTGAAGTCAACTGGCATAGACCAATA	2389
QY	343	LysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIle	362
DB	2390	AAACCAAAATGACTGATATTTACGCACTACTCCACATGATGATTGCAAACTGAGACAG	2449
QY	363	PheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluValAspCysThr	382
DB	2450	CAGGACTACGAACCGCTCAATCTGTATCAATGGAATAATGGATCTTCACCAACTGTACA	2509
QY	383	GluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThr	402
DB	2510	AGCAAAACATTC-----TGCCCTCTCCACGCGAGATTCCAAAATACCAAGTATGAA	2563
QY	403	ThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsnTyrLeu	422
DB	2564	ACCACCGTGAATPACTGTGATGAGAGAAATATATCTTCTTCCCAACACATTAACCTA	2623
QY	423	LeuProGluAlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeuProArgCys	442
DB	2624	ACTCAGGACTCAGAAAGAAATGCTGTCAAAGATGGAAGGTGGCAGTCAATACCTGCTGC	2683
QY	443	ValGlu-----	444
DB	2684	ATTGAATAAATTCATGTTCCAGCCCTTACAAATAGAATGATCTATTATTATCC	2743
QY	444	-----	444
DB	2744	AGATCTTCAGAGAAAGAGAGATTCCATTGAGTCCAGCAGCATGAACTGGAAGCTACA	2803
QY	444	-----	444
DB	2804	TTACAGTATGCTGTGATGATGTTTCAAGATACCTGAGAAATAATAGATACTGCTAC	2863
QY	444	-----	444
DB	2864	ATGGAGAAATGAGACACTCCACCTGCTGTGTGACATTCCTTTGGAGACTCCACCTTCA	2923
QY	444	-----	444
DB	2924	ATTCTCTGCTGACTGTTCTCTTCAAGCTAGAGAGTTACCAACATGGGGAAGGTTACA	2983
QY	444	-----	444



Db 2984 TACCATTGTTCTACAGGCTTTGGAAATGATGACACGACATTTATATATATGCGAAGAGGA 3043  
 444 ----- 444  
 Db 3044 AAGTGTCTGACCCACCAAAATGCATTAAGAGGATGTGACGTTTACCACAGTTAA 3103  
 444 ----- 444  
 Db 3104 AATGCCATAATAGAGAAAGCAAAATCATATAGAGACAGAGACAGACATTC 3163  
 444 ----- 444  
 Db 3164 AGATGCAATCTCTTATCAATGAATGCTCAGACACTGTGACATGTATATAGTCGG 3223  
 444 ----- 444  
 Db 3224 TGGATTGACAGCCAGCTATGCAAAAGATTAATCTCTGTGTGATCCACCATGTGCCAAT 3283  
 444 ----- 444  
 Db 3284 GCTACTATAGTACAGACCAAGAAATATATCTACATGTGTGACAGAGTACGTATGA 3343  
 444 ----- 444  
 Db 3344 TGTAAATAACTTTGGAAGTATTTGGGCAAGTGAAGTATGTGTGAAATGGGATATG 3403  
 444 ----- 444  
 Db 3404 ACAGAAAAACCAAGTCCGAGCTCTGTATTCGACTTGACTTCAACCTTCAAATGTT 3463  
 445 ----- SerThralatyrCysglyProProSerIleasnnglyAspThr 460  
 3464 TTCTCCTTAGCTACACAGGAAATGTGGGCTCTCCACCTATTCACATGAGACATC 3523  
 461 ThrSerPheProLeuSerValTyrProProglySerThralThrValTyrTyrArgCysGlnSer 480  
 3524 ACCTCCCTGTGATACAGATATGACATTCATTCATCACTGCAATATCATATCCAGAAC 3583  
 481 PheTyrLeuGlnGlySerValThrValThrCysArgAsnGlySerIleProSerGluPro 500  
 3584 TATATCTCTTAGAGGAAAGACATATACATGACATGACAAATGCAAGTGTGTGACCA 3643  
 501 ProArgCysLeuAspProCysValValSerGluGlnAsnMetAsnGlySerAsnGln 520  
 3644 CCACATGCTTACATGATGTAAATACAGAAACATTAATGATACACATATAAT 3703  
 521 LeuLysTyrArgAsnAspLysLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540  
 3704 CTCAAATGAGACACACTGAAAGATTTATTCCTTCAGGCGAGATATGGAATTTGA 3763  
 541 CysLysPheProHisLysAlaMetIleSerSerProPheArgAlaIleCysGlnGlu 560  
 3764 TGTAAATATGATATATTAAGCAAGAGATTCACCCGCTTTGCTACAAAGTCATTAAT 3823  
 561 GlyLysPheGluTyrProIleCys 568  
 3824 GGCACCATCAATATCCCACTTGT 3847

RESULT 11  
 A91885 LOCUS A91885 2715 bp DNA linear PAT 22-JAN-2000  
 DEFINITION Sequence 3 from Patent WO9823638.  
 ACCESSION A91885  
 VERSION A91885.1 GI:6740762  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 2715)  
 AUTHORS Schwaible, W. and Sim, R.B.  
 TITLE COMPLEMENT INHIBITOR  
 JOURNAL Patent: WO 9823638-A 3 04-JUN-1998;

FEATURES SCHWAIBLE WILHELM (GB); UNIV LEICESTER (GB)  
 Location/Qualifiers  
 source 1..2715  
 /organism="unidentified"  
 /db\_xref="taxon:3264"  
 BASE COUNT 863 a 518 c 618 g 716 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,79e-107 Length: 2715  
 Score: 1307.00 Matches: 252  
 Percent Similarity: 47.91% Conservative: 69  
 Best Local Similarity: 37.61% Mismatches: 157  
 Query Match: 41.06% Indels: 192  
 DB: Gaps: 7  
 US-09-805-337a-2 (1-569) x A91885 (1-2715)  
 QY 23 CysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsnProphe 42  
 660 TGTGATTTTCCACAAATCAACATGAGACGCTGTATATGAAGAAAGCGGAGACCTAC 719  
 QY 43 SerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheValSerPro 62  
 720 TTCCAGTACCTTAAAGAGAGTACAGTATATGACAAACGGGTTTCAACGCGCT 779  
 QY 63 SerLysSerPheThrPheArgIleThrCysThrGluGlnGlyTyrSerProThrProLys 82  
 780 TCACAGTCTACTAGGACCTACCTCTTGCACAGTAAAGGGTGGAGCCTGAAGTTCCA 839  
 QY 83 CysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerIleGlu 102  
 840 TCCTCAGCGCAAGTATTTTCCATATGTGGAATATGGAATCTTCATACAGCAAGA 899  
 QY 103 IleHisLeuGlnGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGlnAsn 122  
 900 AGATATATAGAGGTGAGTCAAGTCAAAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 959  
 QY 123 AsnGluLysAsnIleSerCysValGluArgGlyTyrSerThrProProIleCys 140  
 960 GTCAAGATATCATATATATTTACAGAGAAATGCTGCTCCCTCCCAATCGTCGGT 1019  
 QY 140 ----- 140  
 1020 ATCAGACTGTGTCATCATCATATATGAAATGAAATGCTTTTCTGATCTGAT 1079  
 QY 140 ----- 140  
 1080 TATACATATGCTCTAAATAGAAACACAGCTATATGATGAACAGGATATGTAACAAT 1139  
 QY 140 ----- 140  
 1140 ACCGAGAAATATACAGAAATATCTTGTCTTCAAGATGATGTCACCTCCACCTCA 1199  
 QY 140 ----- 140  
 1200 TGCATTAGCTGTGATATGCTGTATTTGACAAATTCATAGACTAAGANTATATACACA 1259  
 QY 140 ----- 140  
 1260 TGGTTTAACTCAATGACAAATTAAGCTATGATGTCACATTTGATGAAATGAAATAT 1319  
 QY 141 SerPheThrLysGly----- 145  
 1320 AAACATACCAAGGCTCTATACATGCTACTTATGATGAGATGCTAGTACACCTCTCTGT 1379  
 QY 146 -----GluCysHisValProIleLeuGluAlaAsnValAspAlaGlnProLysLys 162  
 1380 TATGAAAGAAATGACAGCATTTCCCTGTACACCAAGACTTATGTTTTCACAGAGAA 1439  
 QY 163 GluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIleArgVal 182  
 1440 GTAAATATCAAAAGTTGAGATTCGTTGAGTTCTCTTCCGCTTCAGCA---CACAGAGTT 1496

OY	183	GLYSERASPERVALGICYSYRYSINPHEGILYTPSPCTPROASPHERTHCYSYS	202
DB	1497	GGAGCAGATTAGTGCATTCCTACCACCTTTGATGATGCCCTAATTTCCACACGGTCAA	15565
OY	203	GLYGLVALATGSERCYSGLYPROBROGINLEUSERASNGLYGLVALYSGLULLE	222
DB	1557	GGCCAAAGTAAATCTGTGACCACCACTCTTGAAATCCCGCAATGGGGAATTAAGGAACA	1618
OY	223	ARGYSGLUGLUTYGLYHISASNCULVALGILUTYARPCYSASINPROASINPHELLE	242
DB	1617	AAAAAAGTTGATACAGCCATGCTACGCTGGGCAATVGTATGGCAACCTGAACTTCTCA	1678
OY	243	ILEASNGILPROLYSILSILEGINCYVALASPGILGILUTYTPHTRPHEPROTHCYS	262
DB	1677	CTGAAGGGGCCCAATAAATCCAGTGTGTGACGGGAAGTGGACAAAGSTTCCGATATGC	1738
OY	263	VALGILGILVALYSLTHCYSGILYTYLLEPROGULEUGLUTYGLYTYVALGILNPRO	282
DB	1737	GTTGAGTGTGAGAGAACATCTGGAGAACCTTCTCTGAACTTGAGCAAGCTCTCTCAAGTAA	1798
OY	283	SERVALPROPTOTYGLINHISGLYVALSERVALGULVALASNCYSARGASNGILUTYALA	302
DB	1797	TCTGTCCCTCCCTACCATCATGATGAGATTCAGTGGAGTTCACTGTGACGAAGAACCCTTCA	1856
OY	303	METILEGLYASASINMETILETHRCYSILEASNGILYLETPTPHGLULEUPROKETHCYS	322
DB	1857	ATGATGTGACATGTCGATGCTTTTCGATVGTAGTGAGAGGTGGAGCCAGCCTTCCTCAATGT	1918
OY	323	VALAETHPHHISGLINLEULYSARGCYSYSILEAAGLYVALASINLIELLSTHLEULEU	342
DB	1917	GTTGACAAACATCAACTGGAGAGGATGAAGGTAAGACCCCGAAGTCACACTGGCATAGATCAATV	1978
OY	343	LYSLEUSERGILYSGULPHEASINHISASINSEARATILEARTYTYARGCYSERASPILE	362
DB	1977	CATCCAAATPAGAAATGAATTAATCATATAACTTTAGTGAGTTACAGATGTAGACAAAG	2038
OY	363	PHEATGYTARGHISERVALCYSILEASNGILYSTPASPINPROGILVALASPCYSYTHR	382
DB	2037	CAGGAGTATGAACATTCATCATCTGCATCATAGGAATGGATGCCGAACCAACTGTACA	2098
OY	383	GLULYSARGLUGLINPHECYSPROBROPROGINLIEPROASINLISGLINMETHR	402
DB	2097	AGCAAAAGA-----TCTGCCCTCTCCCGCGCATTCCAAAAGCCCAAGTAGTTGAA	2158
OY	403	THRTVALASRTYGLINASPGILYULYSVALAVALLEUCYSLYSGULASNTYLEU	422
DB	2151	ACCACCGGAATACTTGATGGATGGAAAGAAATGATCTGCTTGGCAAGATGGTTAACTTA	2210
OY	423	LEUPROGILUALYSGULILEVALCYSLYSASPGILYATGTPGLINSELEUPROARGYS	442
DB	2211	ACTCAGGGCCCAAGAAATGCTGTGTAAACATGGAAGGTGGCACTCTTACACAGCTGC	2270
OY	442	-----	442
DB	2271	ACGAAAAAATTCATGTTCCAGCCCTAAAATTGAACATGATCTATTAAAGCGCC	2330
OY	443	-----VALGILUSERTHRALATY-----	448
DB	2331	AGTCTCTCAGAAAGAGAGAGATTAAATTAGCTCACCACTTATGAACACGGAATACA	2390
OY	448	-----	448
DB	2391	TTCAGCTATTGCTGTAGAGATGGAATTCAAGATATCTGAGAAATAGGCTAACCTGCAC	2450
OY	449	-----CYSGLYPROBROPSER-----	454
DB	2451	ATGGGAAATGAGACTCTCTGCTCTGTTGTGGAAATACCTGTGGAGCCGCCCACTTCA	2510
OY	455	ILEASINAGLYASPTPHTRSERPHEROLEUSERVALYTYPROPIOLYSETHTHVAL	474
DB	2511	ATTCTCTTGGT---ATTGTTCTGTGACACTAGCAAACTATGACCAATATGAGAGAGAGTT	2567

Oy		475	ThryTArGcYgSlInSerPheTyrlLysleuGlnGlYSerValIshrValThrCysAArgSan 494
Dd		2568	ACATCAATATTGTTCGACAGCCTTTGGAATTTGATGCACAGCATTTATTAAATGTGTAGCA 262
Oy		495	LysGlnTrpSerglInProBrarGcSLeu 504
Dd		2628	GGACAGTGTCTGCAACTCCCAATAATGCATA 2657
RESULT 12			
HUMCHP LOCUS	HUMCHP	2132 bp	mRNA linear PRI 01-NOV-199
DEFINITION	Human complement H factor mRNA, complete cds.		
ACCESION	M17517		
VERSION	M17517.1 GI:180497		
KEYWORDS	H factor; complement; glycoprotein.		
SOURCE	Human liver, cDNA to mRNA, clone R2a.		
ORGANISM	Homo sapiens		
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Day A.J., Ripoché,J., Lyons,A., McIntosh,B., Harris,T.J. and Sim,R.B.		
TITLE	Sequence analysis of a cDNA clone encoding the C-terminal end of human complement factor H (bases 1 to 2132)		
JOURNAL MEDLINE	BioSci. Rep. 7 (3), 201-207 (1987)		
PUBMED FEATURES	88025472 2889480		
source	Location/Qualifiers		
	1..2132	/organism="Homo sapiens"	
	/db_xref="taxon:9606"		
	/map="1q32"		
	1..2132	/gene="HF"	
	<1..1976	/note="HP"	
	/gene="HF"		
	/note="complement H factor"		
	//codon_start=2		
	/protein_id="AAAS2016.1"		
	/db_xref="GI:180498"		
	/db_xref="DB:000-120-041"		
	/translation="DPRNVDPBRKRDQYKGVGLFKFSCKPGFTIVGPSVOCYHGFLGS DLPIKEOVOSGGPELLNGAVKREKTESEYSHSEVEEYCNPRLMKGPNKLOCVD EMTLTPVCIVFESTCGCDIPLEHGMNAOLSSPEYYGDSVENFCSTSPFMIGHSITC HCWMTLPQCVADIMLKCKCSNLIILEHLNKKKEFNMSIRBCRKBEMIHVV INGRMDPEVNSMAOIQLCPPEPPQILPNSNNMTTLTYLRGSEVSYLCQNTILOEE ITCKGRMOSIPLVEKILPCSQPOLIEHOTINSSSDSVIAHGKILYSTCEGFRI ENETNYCYMKWSPPQCQELKDCLSLPFENAIIPMGEEKDYVKAEOVYTATAYRK GPALIKALGEKWSHPSPCIKITDCLSLPFENAIIPMGEEKDYVKAEOVYTATAYRK DGASVTWCINSMTGRPTCRDCSNVPVQNAVIVSROMSKPSGSEBRYOCRSBPY MGEDEVCMICNMWTREPPOCKDKSTGCKPPIDNGDITSPLSVAPASVEYOOCO LYOLEGNKRRTICRNOWSPPKCLHPPVSPREIMENYNALMLWTAKOKILYSTGESV FYCKRGTRLSRSHTLRITTCCKDGKLEYPICAKR"		
BASE COUNT	714 a 385 c 454 g 579 t		
ORIGIN	92 bp upstream of Avail site; chromosome 1q32.		
Alignment Scores:			
Pred. No.:	1.51e-105	Length:	2132
Score:	1284.00	Matches:	246
Percent Similarity:	47.21%	Conservative:	59
Best Local Similarity:	38.08%	Mismatches:	103
Query Match:	40.34%	Indels:	238
DB:	9	Gaps:	1
US-09-805-337A-2 (1-566) x HUMCHP (1-2132)			
Oy		161	LysLysGluSerTyrLysValGlnLysPheSerCysArgLysAsnLeuile 180
Dd		26	AAGAAGACACAGTAATAAGATTGGAGAGGTTGTAATTCCTCGCAAACACAGATTACA 85
Oy		181	ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProasnPheProtmr 200
Dd		86	ATAGTGTGACCCAAATATCGGTTCAGTGACCAACCATTTGGATTGTCTTCACCTGCCAATA 145

```

OY 201 CysIsglyGluValArgSerCysGlyProProGluLeuSerAsnGlyGluValLys 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 TGTAAAGGCAAGTACATCATGTGCTCACCCTCTGACCTCCCAATGGAATGTTAAG 205
OY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 GAAAGAAAGAAAGAAATATGACACAGTGAAGTGGTGAATATATTATGCAATCTCTGA 265
OY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeuPro 260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 TTTCTAAAGAGGACCTTAATAAATTCATATGTTGAGAGTGGACACTTATACCA 325
OY 261 ThrCysValGluGluValLysThrCysGlyTyrIleProGluLeuGlyTyrGlyVal 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 GTGTGTATGTGGAGAGATGACCTGTGGACATATACCTGAACCTTGAGCGCGGCC 385
OY 281 GluProSerValProProTyrGluHisGlyValSerValGluValAsnCysArgAsnGlu 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 386 CAGCTTCTCCCTCTTATCTATGATGAGATTCAGTGAATTCATTCATGCTCAGATCA 445
OY 301 TyrIleMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTyrThrGluLeuPro 320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 446 TTTACAAATGATGGACACAGATCAATTCATGCTATTCATGAGATGAGACCACTTCC 505
OY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 506 CAGTGTGGCAATGATAACTTAAAGATGACCAATCATCAATTTAATTTACTTGG 565
OY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 566 GAACATTTAAAAACAAAGAAATTCGATCTAATCTAATGAGTACAGATGAGATGAGA 625
OY 361 AspiePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluValAsp 380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 GGAAGAAAGAGATGATACACACAGCTGCAATTAATGAGATGAGATGAGACCAATGAG 685
OY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 686 TGCTCAATGGCACAATACAAATTAATGAGTCCACCTCCACATATCCCAATCTCACAAT 745
OY 401 MetThrThrThrValAsnTyrGlnAspGlyGluValAlaValLeuCysLysGluAsn 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 746 ATGACAAACCACTAAATTAATCGGGATGAGAGAAAGATATCTGTTCTTGGCAAGAAAT 805
OY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeuPro 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 806 TATCTAATTCAGGAAGGAAAGAAATTCATGCAAAAGATGACAGATGGCAGTCAATACCA 865
OY 441 ArgCysValGlu----- 444
    ||||| |||||
Db 866 CTCGTGTGTAAGAAATTCATGTTCAACACCCTCAGATAGACAGCAACCATTAAT 925
OY 444 ----- 444
Db 926 TCATCCAGGCTCTCACAAGAAAGTATATGACATGGAGCTAAATGAGTTATCTGTGAG 985
OY 444 ----- 444
Db 986 GGTGCTTCAGGATATCTGAAGAAATGAACACATGCTACATGGAAATGAGCTTCT 1045
OY 444 ----- 444
Db 1046 CCACCTCAGTGGAGAGGCTTCTGTAATCTCCACCGAGATTTCTCATGCTGTGTA 1105
OY 444 ----- 444
Db 1106 GCTCATCATGACAGACGTTATCATGTGAGAGAAAGTATGCTACAAATGTTTGAAGT 1165
OY 444 ----- 444
Db 1166 TTTGGAATTCATGGGCTTCGAAATTCGAAATGCTTAGAGAGAAATGCTTCACCCCTCCA 1225

```

```

OY 444 ----- 444
Db 1226 TCATGCATAAACAGATGTTCTCAGTTACTTACCTTACCTTGAATAATGCCATGCCATGSGA 1285
OY 444 ----- 444
Db 1286 GAGAGAGAGATGTGTATTAAGCGGGTGAAGCAAGTCACTTGTGCAACATATTAC 1345
OY 444 ----- 444
Db 1346 AAATGATGAGCCAGTAAATGTAACATGATTAATAGCAGATGACAGAGAGCCCAACA 1405
OY 444 ----- 444
Db 1406 TGCAGAGCACCTCTCTGTGTATTCGCCGCCACAGTACAAATGCTTATATAGTGCAGA 1465
OY 444 ----- 444
Db 1466 CAGATGATTAATATCATCTGCTGAGAGAGTACCTTATCAATGATGAGAGCCCTTATGAA 1525
OY 444 ----- 444
Db 1526 ATGTTGGGATGAGAGAGATGATGTTTAAATGGAACCTGGACGGAACCACTCAATGC 1585
OY 445 -----SerThrAlaTyrCysGlyProProProSerIleAsnGlyAspThrThrSer 462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1586 AAAGATTTCTACAGAAATGTTGGGCCCTCCACCTATTGACAAATGGGACATTACTTCA 1645
OY 463 PheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSerPheTyr 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1646 TTCGGTGTGATGATATGCTTCACCTTATCATGATGATGAGTACCAATGCGCAGACTTGTAT 1705
OY 483 LysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrProSerGluProProArg 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1706 CAACCTGAGGGTAAACCAACCAATTAACATTAACATTAACATTAACATTAACATTAAC 1765
OY 503 CysLeuAspProCysValValSerGluGlnLysAsnMetLysAsnAsnIleGlnLeuLys 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1766 TGCTTACATTCGTTGATATATCCGAGAAATTAATGAGAAATTAATACATACATTAAG 1825
OY 523 TrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGlnCysLys 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1826 TGGACAGCCAAACAAACCTTTATTCGAGAACAGTGAATGAGTGAATTTCTGTGTA 1885
OY 543 PheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluGlyLys 562
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 CGGGATATTCGTTTTCATCAGCTTCTCACACATTCGACACACATGTTGGATGGAAA 1945
OY 563 PheGluTyrProIleCys 568
    ||||| ||||| |||||
Db 1946 CTGAGATTCCAACTTGT 1963

RESULT 13
HONHAB 1264 bp mRNA linear PRI 27-APR-1993
LOCUS Human factor H homologue mRNA, complete cds.
DEFINITION M65293
ACCESSION M65293.1 GI:183764
VERSION 1
KEYWORDS factor H.
SOURCE Human liver, cdna to mRNA, clone pFH.4b.
ORGANISM Homo sapiens
Mammalia; Euteleostomi;
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Estaller,C., Koistinen,V., Schwaebler,W., Dierich,M.P. and
Weiss,E.H.
Cloning of the 1.4-kb mRNA species of human complement factor H
reveals a novel member of the short consensus repeat family related
to the carboxy terminal of the classical 150-kDa molecule
J. Immunol. 146 (9), 3190-3196 (1991)
JOURNAL MEDLINE
PUBMED 91201892
1826708 Location/Qualifiers
FEATURES

```



LOCUS BC016755 1296 bp mRNA linear PRI 09-NOV-2001  
 DEFINITION Homo sapiens, H factor (complement)-like 1, clone MGC:13525  
 IMAGE:3934474, mRNA, complete cds.  
 ACCESSION BC016755  
 VERSION BC016755.1 GI:16876960  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 1296)  
 AUTHORS Strausberg, R.  
 DIRECT SUBMISSION  
 SUBMITTED (05-NOV-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 CONTACT: MGC help desk  
 EMAIL: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 TISSUE Procurement: CLONTECH  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Institute for Systems Biology  
 CONTACT: <http://www.systemsbio.org>  
 contact: amadan@systemsbiology.org  
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha  
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting  
 Cloning distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 19 Row: a Column: 13  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 11321586.  
 FEATURES  
 Source Location/Qualifiers  
 1..1296  
 /organism="Homo sapiens"  
 /db\_xref="locusid:3078"  
 /db\_xref="taxon:9606"  
 /clone="MGC:13525 IMAGE:3934474"  
 /tissue\_type="Skeletal Muscle"  
 /clone\_lib="NIH\_MGC\_81"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 93..1085  
 /codon\_start=1  
 /product="H factor (complement)-like 1"  
 /protein\_id="AAH16755.1"  
 /db\_xref="GI:16876961"  
 /translation="MMLVSVILIRISVSGEATFCDFPKINHGILYDEERYKPFSSQ  
 VPTEGVYVSECVNFMVSPKSFMTIRICTEGMSPTKCLACFPFVNGHSSGQ  
 TLEGDVOICNTGYLONNENISCVKRWGMSPPKCRSDTSCVPPVONAHILS  
 ROMSKYPSGFRVRECRSPYEMNGDEVMKLNKMTPEPPCKDSCGCGPPRPDND  
 ITSPRLSYAPASSVEYOCNLVLECNKRTKNGMSEPKLHPVLSRIMEND  
 NIALRWAKQKILRTGESAEFVCKRGYRULSSRSHLTWTCTMDKLEYPTCAKR"  
 BASE COUNT 440 a 240 c 250 g 366 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3.22e-85 Length: 1296  
 Score: 1054.50 Matches: 216  
 Percent Similarity: 43.49% Conservative: 31  
 Best Local Similarity: 38.03% Mismatches: 80  
 Query Match: 33.13% Indels: 241  
 DB: 9 Gaps: 5  
 US-09-805-337a-2 (1-569) x BC016755 (1-1296)  
 QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
 DB 93 ATGTGGCTCGTGTGACGTAAATCTATATCTACGAGATATCTCTGTTGGGGAGAGACA 152  
 QY 21 ThrLeuCyAspPheProLysIleHisHisGlyPheLeuTrpAspGluAspTyrAsn 40

DB 153 ACATTTTGTGATTTTCCAAAATAACCATGAGATTCATATGATCAAGAAAATATTAAG 212  
 QY 41 ProPheSerGlnValProThrGlnGlyGlnValPheLeuTyrSerCysGluTyrAsnPheVal 60  
 DB 213 CCATTTTCCCAAGGTTCTACAGGGGAAGTTTCTATTAATCTCTGTGAATAAATTTTGTG 272  
 QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGlnGluGlyTyrPseProThr 80  
 DB 273 TCTCTCTTCAAAATATTTTGGACGTCGATACATGACACAGAGAGATGTCACCAACA 332  
 QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlnHisSerGluSerSer 100  
 DB 333 CCAAGAGTGTCCACAGCTGTGTTCTCTCTTGTGAAAATGTCATTCGATTCGATTCACA 392  
 QY 101 GlnLeuIleHisLeuGlnGluGlyAspTrpValGlnIleLeuCysAsnThrGlyTyrSerLeu 120  
 DB 393 GGACAAACACATCTCGAAGGTGATACGTGCAAAATTTTGCACACAGCATACAGACTT 452  
 QY 121 GlnAsnAsnGlnLysAsnIleSerCysValGlnArgGlyTrpSerThrProProIleCys 140  
 DB 453 CAAACCAATGAGACACACATTTTCATGTGTAGAACGGGCTGTGTCACCCCTCCAAATCC 512  
 QY 141 SerPheThrLysGlnGlyCysHisValProIleLeuGlnIleAsnValAspAlaGlnPro 160  
 DB 513 AGGTCCACTGCAC----- 524  
 QY 161 LysLysGlnSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180  
 DB 524 ----- 524  
 QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProThr 200  
 DB 524 ----- 524  
 QY 201 CysLysGlnGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGlnValLys 220  
 DB 525 -----ACTTCCTGTGTGATCCGCCACAGTCAAAATGTCATATGACTG 569  
 QY 221 GlnLeuArgLysGlnGluTyrGlnHisAsnGlnValValGlnTyrAspCysAsnProAsn 240  
 DB 570 TCGAGACAGATGAGTAAATATCCATCTGGTAGAGACACTGTATGAA----- 617  
 QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro 260  
 DB 617 ----- 617  
 QY 261 ThrCysValGlnGlnValLysThrCysGlyTyrIleProGlnLeuGlnTyrGlyVal 280  
 DB 617 ----- 617  
 QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGlnValAsnCysArgAsnGlu 300  
 DB 618 -----TGTAAGAGCCCT 629  
 QY 301 TyrAlaMetIleGlnAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGlnLeuPro 320  
 DB 630 TATGAATGTTTGGGAGTGAAGAGTGTATGCTTAATGGAAGTGCAGAGAA----- 683  
 QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340  
 DB 683 ----- 683  
 QY 341 LeuLeuLysLeuSerGlnLysGlnPheAsnHisAsnSerArgIleArgTyrArgCysSer 360  
 DB 683 ----- 683  
 QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGlnValAsp 380  
 DB 683 ----- 683  
 QY 381 CysThrGlnLysArgGlnGlnPheCysProProProGlnIleProAsnIleGlnAsn 400



```
Db 668 ----- 668
QY 381 CysThrGluysArgGluInPheCysProProProGlnIleProAsnAlaGlnAsn 400
Db 669 ----- 677
      |||||
      CCACCTCA-----
QY 401 MetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
Db 677 ----- 677
QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrrpGlnSerLeuPro 440
Db 678 ----- 686
      |||||
      TCGAAGAT-----
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db 687 ----- 734
      |||||
      -TCACGGGAAATGTGGGCCCCCTCCACCTATTGACAAATGGGACATT 734
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db 735 ACTTCATTCCCGTGTGACGATATGCTCCAGCTTCATCAGTGCAGTACCAATGCCAGAAC 794
QY 481 PheTyrLysLeuGlnIlySerValThrValThrCysArgAsnLysGlnTrrpSerGluPro 500
Db 795 TTGTATCACTTGAGGGTAACAAGCAGAAATGACATGACAAATGGTCAAGAACCA 854
QY 501 ProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
Db 855 CCAAAATGCTTACATCCGTGTGTAATATCCCGAAATTTGCAAAATTTATACATRGCA 914
QY 521 LeuLysTrrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
Db 915 TTAAAGGTGGACCAACCAAGACTTTATTGAGAAACAGGTGAATCAGCTCAATTGTCG 974
QY 541 CysLysPheProHisLysAlaMetIleSerSerProPheArgAlaIleCysGlnGlu 560
Db 975 TGTAAACGGGAGATATGCTCTTTTCATCAGCTTTCACACACATTGGCAACACATGTGGGAT 1034
QY 561 GlyLysPheGluTyrProIleCys 568
Db 1035 GGGAAACTGGAGTATCCAACTGT 1058
```

Search completed: July 23, 2003, 10:49:57  
Job time : 4214 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 23, 2003, 03:04:22 ; Search time 347 Seconds  
(Without alignments)  
3692.758 Million cell updates/sec

Title: US-09-805-337a-2

Perfect score: 3183  
Sequence: 1 MLLFSEVILITSWSTVGSGS.....SSPPFRATCEGKFEYPICE 569

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O=cgq2\_1/USPTO.spool/US09805337/runat\_17072003\_135605\_18885/app\_query.fasta.1.711  
-DB=N\_Geneseq\_101002 -OEMT=fastap -SUFFT=ring -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=b1cs -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdt  
-LIST=45 -DOCMALIGN=200 -THR=SCORE=pcpt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTEFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09805337 @CGN.1.1.263 @runat\_17072003\_135605\_18885 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEIOUDRY -NEG\_SCORES=0 -NAIT -DSPBLOCK=100 -JONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3183	100.0	2823	AAD37114	Human factor H-rel
2	2733	85.9	1707	AAD37115	Human factor H-rel
3	2162	67.9	1225	ABK49924	Human novel polynu
4	1691	53.1	984	ABU57707	Human sbg614126com
5	1599	50.2	1095	ABU57706	Human sbg614126com
6	1303	40.9	2548	ABQ55008	Human ovarian anli
7	1051.5	33.0	1293	AACT7947	Human cancer assoc
8	1050.5	33.0	1266	ABN58888	Gene #2386 used to
9	1001.5	31.5	2033	AAV34737	Human complement F
10	927.5	29.1	1040	ABN87268	Gene #3766 used to
11	756.5	23.8	2177	AAH82402	B subunit of human
12	710	22.3	1269	ABN97281	Gene #3779 used to
13	681.5	21.4	1929	AA506029	Angiotensin conver
14	681	21.4	1315	ABN97342	Gene #3840 used to
15	515	16.2	767	AAV02792	Human partial comp
16	487	15.3	759	AAV39793	Gastric cancer ass
17	466	14.6	688	AAV02794	Clone pBS3FH2910
18	432.5	13.6	590	AAV02795	Clone pBS3FH2910
19	424	13.3	3327	AAQ10989	B lymphocyte membr
20	395	12.4	2929	AA503887	Human secreted pro
21	395	12.4	3566	AAH75787	Receptor 222 codin
22	395	12.4	3706	ABN83421	Human gene GS96663
23	395	12.4	3804	ABAB6548	Human cDNA SEQ ID
24	395	12.4	3804	AA528968	cDNA encoding for
25	395	12.4	3804	AA528968	Human cDNA encodin
26	395	12.4	3804	AA531587	cDNA encoding nove
27	395	12.4	3804	ABK43685	DNA encoding novel
28	395	12.4	3891	ABK13101	Human REPR 12 cDN
29	394	12.4	10878	AAD32025	Human C3b/C4b comp
30	389.5	12.2	1194	AA506052	Angiotensin conver
31	389	12.2	6153	ABAB3880	Human POU11 nucle
32	385	12.1	5420	AAQ11643	Partial human comp
33	385	12.1	5420	AAZ38151	Human C3b/C4b rece
34	385	12.1	5420	ABU58682	Lung cancer relate
35	385	12.1	5420	ABU1637	Human C3b/C4b rece
36	385	12.1	6951	AAH91477	CRI protein DNA.
37	385	12.1	6951	AAQ11642	Entire human compl
38	385	12.1	6951	AAZ38150	Human C3b/C4b rece
39	385	12.1	6951	AAH84738	Human cDNA differe
40	385	12.1	6951	ABAB1636	Human C3b/C4b rece
41	385	12.1	7313	AAI58380	Human polynucleoti
42	385	12.1	7821	AA564474	DNA encoding novel
43	385	12.1	9038	AA564290	DNA encoding novel
44	384.5	12.1	11230	AAD32026	Mouse C3b/C4b comp
45	383.5	12.0	2575	ABQ72635	Human MDT encodin

## ALIGNMENTS

RESULT 1	
AAAD37114	AAAD37114 standard; cDNA: 2823 BP.
AC	AAAD37114:
AC	21-AUG-2002 (first entry)
DT	
XX	Human factor H-related protein (FHR-5) encoding cDNA.
DE	
XX	Human: factor H-related protein; FHR-5; detection: C3b-9 complement;
KW	immunohistochemical; diagnostic; immunological; biopsy; prevention;
KW	gene; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
	CDS 94..1803

FT /tag- a  
XX /product- "Human factor H-related protein (FHR-5)"  
PN MO200168695-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 13-MAR-2001; 2001MO-US07868.  
XX  
PR 13-MAR-2000; 2000US-0188870.  
XX  
PA (BAXT ) BAXTER INT INC.  
XX  
PI Murphy BF;  
XX  
DR WP1; 2001-582437/65.  
XX  
DR P-PSDB; AAE23195.  
XX  
PT New factor-H related protein 5 that binds to complement component C3b,  
PT useful for raising antibodies suitable for detection of C3b-9  
PT complement complexes,  
PS  
PS Claim 3; Page 50-52; 61pp; English.  
XX  
XX The invention relates to a novel human factor H-related protein  
CC (FHR-5), its corresponding nucleic acid and antibodies directed  
CC to it. The antibody is useful for detecting C3b-9 complement  
CC complexes, by contacting a cell or tissue with Ab and detecting  
CC the binding of Ab to the cell or tissue. Ab is useful as  
CC immunohistochemical diagnostic reagent to detect immunological  
CC deposits in biopsied tissues, and also for preventing the  
CC association of FHR-5 with activated complement. The present  
CC sequence is human FHR-5 cDNA.  
XX  
SQ Sequence 2823 BP; 960 A; 531 C; 546 G; 786 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 1,66e-272 Length: 2823  
Score: 3183.00 Matches: 569  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0  
  
US-09-805-337A-2 (1-569) x AAD37114 (1-2823)  
QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
DB 94 ATGTGCTCTTATTGAGTGAATCCCAATCTCATGTGATCCAGCTTGGGGAGAGAGA 153  
QY 21 ThrLeuGlyAspPheProLysIleHisGlyPheLeuTrpAspGluGlyProLys 40  
DB 154 ACCTTTTGATTTCCAAATAACACCATGATTTCTGATGATCAAGAAAGATTATTAAC 213  
QY 41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTrpAsnPro 60  
DB 214 CCTTTTCCCAAGTTCTACAGGGAGATTCTATTACTCCGTGAATATATATTTGTG 273  
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluLysTrpSerPro 80  
DB 274 TGCCTCTCAAAATCCCTTTGAGCTGCATACATGCAGAGAGAGATGGTCCACACA 333  
QY 81 ProLysCysLeuArgNeLcysSerPheProPheValLysAsnGlyHisSerGluSer 100  
DB 334 CCGAAGTGTCTGAGATGTCTCCTTTCTTTGCAAAATGTCATCTGAAATCTTCA 393  
QY 101 GlyLeuIleHisLeuGluLysPheTrpValGlnIleIleCysAsnThrLysTrpSer 120  
DB 394 GGACTAATACATCTGCAAGGTGATACTGTAACAATATTATTGACACAGATACACCTT 453  
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProPheLys 140  
DB 454 CAAACAAATGAGAAATAACATTTCTGTGTGAGAAAGGGGCTGGTCCACTCTCCATATGC 513

QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160  
DB 514 AGCTTCACTAAGAGAAATGTCAATGTCTCCATTTTGAAGCCATGTACATGCTGAGCA 573  
QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeu 180  
DB 574 AAAAAAGAAAGCTACAAAGCTGAGAGCTTGAATTTCTCCGACGAAATAATCTTATA 633  
QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnProPhe 200  
DB 634 AGAGTTGATCAGACTCAGCTCATGTACCAATTTGGGTGGTCACTTACTTCCAAACA 693  
QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysVal 220  
DB 694 TCGAAGAGCAAGTAGCATCATGTGCTCCACCTCCCTCAACTCTCCCAATGTAAGTTAG 753  
QY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240  
DB 754 GAGATTAAGAAAGAGCAATATGACACAAATGAAGTGAATATGATGCAATCTTAAT 813  
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260  
DB 814 TTTATATATAACGGGCTTAGAAATACATGTGTGATGCAATAGCAACACTTACC 873  
QY 261 PheCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyVal 280  
DB 874 ACTTGCTTCAACAACTGAAACATGTGATACATCTGAACTGCAACTCAGATCGTATGTT 933  
QY 281 GlnProSerValProProTrpGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300  
DB 934 CAGCCGCTGCTCCCTCCATCAACATGAGTTTCACTGAGTGAATTCGCAAAATGAA 993  
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320  
DB 994 TATGCAATGATTCGAATATACATGATTACTGTATTAAAGAAATAGCACAGACTTCT 1053  
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340  
DB 1054 ATGTGTTGTGCAACACACCACTTAAGAGTGCAGAAATGAGGATTAATTAATAAACA 1113  
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360  
DB 1114 TTACTCAAGCTATCTGGGAGAGATTAATCAATTAATCTAGAAATGCTTACAGATGTTCA 1173  
QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380  
DB 1174 GACATCTTCAGATPACAGGCATCAGTCTGTATTAACGGGAATGGAATCTCAAGTAGAC 1233  
QY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400  
DB 1234 TGCACGAAAAAAGGCAAAATTCGCCACGCCCTCCAGATACCTATATCTCAGAAAT 1293  
QY 401 MetThrThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420  
DB 1294 ATGACACACACAGTGAATATATCAGATGAGAAAGAAAGTACGTCTCTGTAAAGAAAC 1353  
QY 421 TyrLeuLeuProGluLysGluIleValCysLysAspGlyArgTrpLysSerLeuPro 440  
DB 1354 TATCTACTTCCAGAAAGCAAAAGAAATTTATATGAAGATGAGACGATCATTATACA 1413  
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460  
DB 1414 CGCTGTGTGAGTCTACTCATATGTTGGGCCCTCCATCTATTAACAAATGAGATATAC 1473  
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480  
DB 1474 ACCCTATTCCTCATATACATATATCTCCAGGGGTGAACGTAACGTAACGTAACGTAAC 1533  
QY 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500  
DB 1534 TTCTATAAACTCCAGGGCTCTGTACTATCAATCGAGAAATTAACAGTGTGTGACAGACA 1593

QY	501	ProatgCysLeuAspProCysValValSerGluGluIuAsnMetAspLysAsnAsnIleGln	520
Db	1594	CCAAKATCCCTAGATCCATGATGTTGTAATCTGAGAAACATGAACAAAATATACATACAG	165
QY	521	LeuLysTPAPcAsnAspGlyLysLeuTyralaLysThrGlyAspAlaValGluPheGln	540
Db	1654	TTAAATATGCAAAACGATGCAAAATCTATATGCAAAAACAGGGGATGCTTGTGATTTCCAG	171
QY	541	CysLysPheProHisLysAlaMetIleSerSerProPheArgAlaIleCysGlnGlu	560
Db	1714	TGTAATATCCACATAAAGCGCATATTCATCACCACCATTTGAGGCAATCTGTGAGGAA	177
QY	561	GlyLysPheGluTyProIleCysGlu	569
Db	1774	GGGAATTTGAATATCATCTATATGTGAA	1800

RESULT 2  
AAD37115

ID	AAD37115	standard; DNA; 1707 BP.
AC	AAD37115;	
XX		
DT	21-AUG-2002	(first entry)
XX		
DE	Human factor H-related protein (FHR-5) degenerate DNA.	
XX		
KW	Human; factor H-related protein; FHR-5; detection; C5b-9 complement;	
KM	immunohistochemical; diagnostic; immunological; biopsy; prevention;	
KM	gene; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200168695-A2.	
XX		
PD	20-SEP-2001.	
XX		
PE	13-MAR-2001; 2001WO-US07868.	
XX		
PR	13-MAR-2000; 2000US-0188870.	
PA	(BAXT ) BAXTER INT INC.	
XX		
PI	Murphy BF;	
XX		
DR	WPI; 2001-582437/65.	
XX		
PT	New factor-H related protein 5 that binds to complement component C3b,	
PT	useful for raising antibodies suitable for detection of C5b-9	
PT	complement complexes.	
XX		
PS	Claim 3; Page 54-55; 61pp; English.	
XX		
CC	The invention relates to a novel human factor H-related protein	
CC	(FHR-5), its corresponding nucleic acid and antibodies directed	
CC	to it. The antibody is useful for detecting C5b-9 complement	
CC	complexes, by contacting a cell or tissue with Ab and detecting	
CC	the binding of Ab to the cell or tissue. Ab is useful as	
CC	immunohistochemical diagnostic reagent to detect immunological	
CC	deposits in biopsied tissues, and also for preventing the	
CC	association of FHR-5 with activated complement. The present	
CC	sequence is human FHR-5 degenerate DNA.	
XX		
SO	Sequence 1707 BP; 340 A; 170 C; 280 G; 231 T; 686 other;	

Alignment Scores:

Pred. No.:	7,51e-233	Length:	1707
Score:	2733.00	Matches:	475
Percent Similarity:	83.48%	Conservative:	0
Best Local Similarity:	83.48%	Mismatches:	94
Query Match:	85.86%	Indels:	0
DB:	23	Gaps:	0

US-09-805-337A-2(1-569) x AAD37115 (1-1707)

QY	1	MelleuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly	20
Db	1	ATGTTNTNTNTNTNTNTGCTNATHTYTNATHTSNNGGTGNTSMNACNGTNGGNGCAGCGN	60
QY	21	ThrlaucysAspPheProLysIleHisIleSdLpheuTrpAspGluGluAsnTrpAsn	40
Db	61	ACNTNTGTCGATTTCCNARATACATCAAGNTTTTNTATACATGARGARATATAAT	120
QY	41	ProPheSerGluValProThrGlyGluValPheTrpTrpSerGlyTrpAsnPheVal	60
Db	121	CCNTTWTMNCARGTNCNCACNGCAGAROTNTTATATAMVSNMTGYGARTAAATVTGNT	180
QY	61	SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr	80
Db	181	WSNCCNMSNAARMSNTTGTGACNMGNATHACNTGACNGARGARGDNTGGMSNACN	240
QY	81	ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer	100
Db	241	CCNARCTGTYTNMGNTATGTCGWTSMSTTTCNTTGTNTNARAAYGGNCATVMSGAMNSMN	300
QY	101	GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu	120
Db	301	GGNTTNTATHCAYTNGCARGCGNGAACNCTNCARATHTHTGCAAYACNGGTATWSNYTN	360
QY	121	GlnAsnAsnGlyLysAsnIleSerCysValGluArgGlyTrpSerTrpProIleCys	140
Db	361	CARATAYAYGARRAARAATATMTSNTGCTGTCNARBNNGSNTGGMSNACNCCNATHTCY	420
QY	141	SerPheThrLysGlyGlyGlyCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro	160
Db	421	WSNTTTCACNARCGCATCTGCATGCTCNCATHTTNGARGCNAATGTNGAAGCCNACRCN	480
QY	161	LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle	180
Db	481	AARARARGRMSNTATYAARGTNGCAGTCYNTNTNARITTYMSNTGTMGNARAARATYTNATH	540
QY	181	ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr	200
Db	541	MGNGTNGSMNGSNGAYWSNGTNCARTGTATYACRTTGTGNTGMSNCCNAATVTTCNACN	600
QY	201	CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys	220
Db	601	TGTARAGGNCARGTMTGMSNTGTGGGNCNCCNCCNARITTNMSNAATGGCAGACTNAR	660
QY	221	GluIleArgLysGluGlnLysGlyHisAsnGluValGlyTrpAspCysAsnProAsn	240
Db	661	GARATHMGNAARGARATATGGNCATVAYGARTGNTNGARTRATYGATVGTAAACCNAA	720
QY	241	PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro	260
Db	721	TTTATTAHAAHYGGNCCNAARAARATHCARGTGNTNGAYGGNGARGTGACNACNYTNCCN	780
QY	261	ThrCysValGluGlnValLysThrCysGlyTrpIleProGluLeuGluTrpGlyTrpVal	280
Db	781	ACNTGTCGNGARCARGTNAARACNTGYGANTATATGCCNGARITNGARTATAGGTAATG	840
QY	281	GlnProSerValProProTrpGlnHisCysLysSerValGluValAsnCysArgAsnGlu	300
Db	841	CARCCNMSNGTNCNCCNTATVACARCATYAGSNTGNTMSNGTNGARGTNAATVGTGMSNAAGAR	900
QY	301	TyrAlaMetCileGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro	320
Db	901	TATGCNATGATGAGNAATATATGCTHACNTGNTATHAAYGGNATHHTGGACNGARTTMCN	960
QY	321	MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr	340
Db	961	ATGTCGTGNCNACATCAVCARTYTNARBMGNTGYAARATHGCGNGGNTAAATATHAARCN	1020
QY	341	LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer	360
Db	1021	YTNTNTNARTNMSGCAAGARTTAAATCAATVAYWSNMGNTATATAGTNTATGWSN	1080

QY 361 AspllepeaTyrArgHisSerValCysIleasnGlyLysTrpAsnProGluValAsp 380  
DB 1081 GAYATHTTYMGNTATMGCAAYMSNGTNTGATHAAYGGNAARTGGAAYCCNARGSTNGAY 1140  
QY 381 CysTrnGluLysArgGluGlnPheCysProProProGlnIleProAsnIleGlnAsn 400  
DB 1141 TGTACNGAAMAAGCARGARTTYTGCCNCCNCCCAATTCNNAAATCCNARAY 1200  
QY 401 MetTrnThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420  
DB 1201 AGACACACACAGTAAATATATACAGATGAGNARARAGTNGCTATNTGTATACARARAY 1260  
QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440  
DB 1261 TATYNTNYCCNGACGACNARARATHTGTATGATGAGAYGGMGNTGGCARGMSNTYTCN 1320  
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460  
DB 1321 MONTGTGNGARMSNACNCTATATGTGNCNCCNCCNCAATTAATATATGNGATACN 1380  
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480  
DB 1381 ACNMGNTTYCCNYTMSNGTNTATACCCNCCNMGSNACNGTACNTATATATGNTGCARMSN 1440  
QY 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500  
DB 1441 TTYTATATATTCAGAGGMSNCTNACNCTATGAGNAAATCAATCGMSNARCCN 1500  
QY 501 ProArgCysLeuAspProCysValAlaSerGluGluAsnMetAsnLysAsnIleGln 520  
DB 1501 CCNMGNTTYCCNYTMSNGTNTATACCCNCCNMGSNACNGTACNTATATATGNTGCARMSN 1560  
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540  
DB 1561 YTNARTGAGGMAAGCAGGNAARNTATATGCGNARACNGGAGCNGTNGARTTCAR 1620  
QY 541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560  
DB 1621 TGTATATTCCTCCATATAGCCNATGATHTMSNMSNCCNCTTYMGNCNATHTGTCAR 1680  
QY 561 GlyLysPheGluTyrProIleCysGlu 569  
DB 1681 GGNARTTYGARTATCCNATHTGTCAR 1707  
RESULT 3  
ABK94924  
ID ABK94924 standard; cDNA; 1225 BP.  
AC  
XX  
XX  
DT 30-AUG-2002 (first entry)  
DE  
XX  
XX  
Human novel polynucleotide #35.  
KM  
XX  
Human: gene; ss: inflammatory condition; shock; sepsis; immune response;  
cancer; wound healing; central nervous system disease; hematopoiesis;  
peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
bone degenerative disorder; periodontal disease; reperfusion injury;  
lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
allergic condition; thrombolysis; thrombosis; coagulation disorder;  
fungal infection.  
KM  
XX  
XX  
Homo sapiens.  
OS  
XX  
XX  
W0200244340-A2.  
PN  
XX  
XX  
06-JUN-2002.  
PD  
XX  
XX  
30-NOV-2001; 2001WO-US47004.  
PF  
XX  
XX  
30-NOV-2000; 2000US-0028952.

XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YF, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
XX  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX  
DR WPI: 2002-508509/54.  
XX  
DR P-PSDB: AB666700.  
XX  
PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing  
PS  
PS Claim 1: Page 406-407; 672pp; English.  
CC The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischemia-reperfusion injury, shock, sepsis, immune responses  
CC and cancer and for promoting wound healing. The sequences are used to  
CC induce the proliferation of neural cells and regeneration of nerve and  
CC brain tissue, and are useful for the treatment of central and peripheral  
CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
CC cell disorders and platelet disorders such as thrombocytopenia,  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
CC disease. The sequences of the invention are also useful for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, immune deficiencies and disorders  
CC including severe combined immunodeficiency (SCID), bacterial or fungal  
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
CC novel polynucleotides of the invention.  
XX  
SQ Sequence 1225 BP; 425 A; 243 C; 243 G; 314 T; 0 other;  
Alignment Scores:  
Score: 2,4e-182 Length: 1225  
P-Id: 2162.00 Matches: 384  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 67.92% Indels: 0  
DB: 24 Gaps: 0  
US-09-805-337A-2 (1-569) x ABK94924 (1-1225)  
QY 186 SerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProThrCysLysGluVal 205  
DB 24 TCAGTCAATGTTACCAATTTGGGTGCTACCTTAATCTCCACATGCAAGCAACAGTA 83  
QY 206 ArgSerCysGlyProProProGlnLeuSerAsnGlyValLysGluIleArgLysGlu 225  
DB 84 CGATCATGTGTGCCACCTCCCACTCCCACTGATGTAAGGATTAAGAGATTAAGAAACAG 143  
QY 226 GluTyrGlnLysAsnGluValValGluTyrAspCysAsnProAsnPheIleAsnGly 245  
DB 144 GAATATGACACACATGAGATGATGATATATATGCAATCTTAATTTATATATACGG 203  
QY 246 ProLysLysIleGlnCysValAspGlyLysTrpThrIleAsnProThrCysValGluGln 265  
DB 204 CCTAAGAAATATACATGCGTGCATGAGACATGACCACTTACCACTGTTGACACAA 263  
QY 266 ValLysThrCysGlyTyrIleProGluLeuGluLysGlyTyrValGlnProSerValPro 285  
DB 264 GTGAATACATGTTGATACATACCTGAACTGAGATGAGGTATATGTTACGCTGCTCCT 323  
QY 286 ProTyrGlnHisGlyValSerValGluValAsnGlyCysArgAsnGlyLysTrpAlaMetIleGly 305

```

Db      324 CCTATCAACATGAGCTTCACTCGAGGTGAATTCAGAAATATGCAATGATTCGA 383
Oy      306 AsnaSmcIleThrCysIleasnGlyIleTrpHgluLeuProMetCysValAlaThr 325
Db      304 AATACATGATTACTGTTATTTATGTAATATGACAGACCTTCTTATGTTGTTGCACA 443
Oy      326 HisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThrLeuLeuLysLeuSer 345
Db      444 CACCAACTTAAGAGGTGCAAAATACAGAGATTAAATATAAAACAATTACTCAAGCATCT 503
Oy      346 GlyLysGluPheAsnHisAsnSerArgIleArgIleArgCysSerAspIlePheArgTyr 365
Db      504 GGGAAAGAAATTTAAATCAATTAATCTGAAATACGTTACAGATGTTGACACATCTTCAGATAC 563
Oy      366 ArgHisSerValCysIleasnGlyLysTrpAsnProGluValAspCysThrGluLysArg 385
Db      504 AGGCATCTAGTGTGTTATTAACGGGAAATGGATCCTGAGATGACTGCAAGAAAAGAG 623
Oy      386 GluGlnPheCysProProProGlnIleProAsnAlaGlnAsnMetThrThrVal 405
Db      624 GAACAATCTGCGCCACCGCCACCTCAGATACCTAATGCTCAGAAATATGACAAACACAGT 683
Oy      406 AsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsnTyrLeuLeuProGlu 425
Db      664 AATTATCGAGATGGAGAAAAGTACCTGTTCTCTGTTAAAGAAACCTATCTACTCCAGAA 743
Oy      426 AlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeuProArgCysValGluSer 445
Db      744 GCAAAAGAAATTTGTTATGTTAAAGATGACAGATGCAATTCATTACCGCTGTTGACTCT 803
Oy      446 ThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThrThrSerPheProLeu 465
Db      804 ACTGCATATGTGGGCCCCCTCATCTATTACAAATGAGATACCACTCATCTCCATTA 863
Oy      466 SerValTyrProProGlySerThrValThrTyrArgCysGlnSerPheTyrLysLeuGln 485
Db      864 TCAGTATATCCCTCCAGGCTCAACAGTACGCTGCTGCCAGTCCCTCTATTAACCTCCAG 923
Oy      486 GlySerValThrValThrCysArgAsnLysGlnTrpSerGluProProArgCysLeuAsp 505
Db      924 GGCTCTGTTACTGTTAAATCGCAAAATTAACACTGGTTCGAACCCACCAAGATGCTTAGAT 983
Oy      506 ProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGlnLeuLysTrpArgAsn 525
Db      984 CCATGTGTCGATCTGAAGAAAACATGAACAAAATAACATACAGTTAAATGGAGAAAC 1043
Oy      526 AspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGlnCysLysPheProHis 545
Db      1044 GATGGAAAACCTTATGCAAAAACAGGGGATGCTGTTGAATTCAGTAAATTCCTCCACAT 1103
Oy      546 LysAlaMetIleSerSerProProPheArgAlaIleCysGlnGluGlyLysPheGluTyr 565
Db      1104 AAAGGAGATATTCATCAACACCATTTTCAGCAATCTGCAGAGAGGAAATTTGAATAT 1163
Oy      566 ProIleCysGlu 569
Db      1164 CCTATATGTGAA 1175

```

## RESULT 4

ABL57707 standard, DNA; 984 BP.

ABL57707;

08-OCT-2002 (first entry)

Human sbg614126complfH gene #2.

Human; secreted protein; immunosuppressive; cytostatic; neutrotropic;  
 neuroprotective; antitumor; vulnerrary; antineoplastic; ophthalmological;  
 antiparkinsonian; antileptemic; antiatherosclerotic; dermatological;  
 hypotensive; cerebroprotective; vitucide; antinflammatory; diabetes;  
 malignant tumor; hypertension; hypotension; obesity; bulimia; anorexia;

```

KW      asthma; manic depression; dementia; delirium; mental retardation;
KW      Huntington's disease; Tourette's syndrome; schizophrenia;
KW      mental disorder; sexual development disorder; blood cascade dysfunction;
KW      stroke; growth disorder; gene; ds.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      CDS      1..984
FT      /tag= a
FT      /product= "sbg614126complfH"

```

WO200222802-A1.

21-MAR-2002.

13-SEP-2001: 2001WO-US28462.

13-SEP-2000: 2000US-232455P.

13-SEP-2000: 2000US-232463P.

02-OCT-2000: 2000US-237283P.

07-NOV-2000: 2000US-246269P.

20-NOV-2000: 2000US-252049P.

(SMK ) SMITHKLINE BECHAM CORP.

(SMK ) SMITHKLINE BECHAM PLC.

(GLAX ) GLAXO GROUP LTD.

Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smuth RF, Strum JC;  
 Xie Q;  
 WPI: 2002-393963/42.  
 P-PSDB; ABB80571.

Novel isolated secreted polypeptides and polynucleotides encoding them  
 useful for treating cancer, Alzheimer's disease, tumor metastasis,  
 autosomal recessive atypical hemolytic uremic syndrome, wound healing  
 disorder

2: Page166: 246pp: English.

The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)  
 (secreted polypeptide) which is encoded by any one of 44 polynucleotide  
 sequences (ABL57705-ABL57748) given in the specification. The  
 CC polypeptides have immunosuppressive, cytostatic, neutrotropic,  
 CC neuroprotective, antitumor, vulnerrary, antineoplastic, ophthalmological,  
 CC antiparkinsonian, antileptemic, antiatherosclerotic, dermatological,  
 CC hypotensive, cerebroprotective, vitucide, and antinflammatory activity.  
 CC The polynucleotide and polypeptide are useful for treating diabetes,  
 CC malignant tumours, hyper- and hypotension, obesity, bulimia, anorexia,  
 CC asthma, manic depression, dementia, delirium, mental retardation,  
 CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental  
 CC or sexual development disorders, and dysfunctions of the blood cascade  
 CC system including those leading to stroke.

Sequence 984 BP: 309 A; 201 C; 194 G; 280 T; 0 other;

Alignment Scores:

pred. No.:	1..18e-140	Length:	984
Score:	1691.00	Matches:	326
Percent Similarity:	57.47%	Conservative:	1
Best Local Similarity:	57.29%	Mismatches:	0
Query Match:	53.13%	Indels:	242
DB:	24	Gaps:	2

US-09-805-337A-2 (1-569) x ABL57707 (1-984)

1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGluGly 20  
 1 ARGTTGCTCTTATTTCAGTATATCCAAATCTCATGGTATCCACACTTGGGGAGAGAGA 60  
 21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsn 40

```

Db      61 ACACCTTGGAATTTCCAAAAATACACCATGATTTCTGTATGATGACAGATTATTAAC 120
QY      41 ProPheSerGlnValProThrGlnGluVal1PheTYrTyrSerCysGluTyrAsn1PheVal 60
Db      121 CTTTTTCCCAAGTCTCTACAGGGAGATTCTTATCTACTCCGTAATATATATTTTGTG 180
QY      61 SerProSerLysSerPheThrPheArgIleThrCysTyrGlnGluGlu1YrPseProThr 80
Db      181 TCTCCCTCAAAATCCCTTTGACTCCCATACATCCACAGAGAGAGATGTCACACA 240
QY      81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100
Db      241 CCGAATGCTCTCAGATGTGTTCTTCTTCTTGTAAAAAGTGTATTTCTGATCTTCA 300
QY      101 GlyLeuIleHisLeuGlnGluLysPheValGlnIleIleCysAsnThrGlyTyrSerLeu 120
Db      301 GCACATATACATCTGCAAGGTGATCTGATACAAATATATTGCAACACAGATATACACCTT 360
QY      121 GlnAsnAsnGlnLysAsnIleSerCysValGluArgGlyTyrPseThrProPhe1Cys 140
Db      361 CAAACAAATGAGAAAAACATTTCTGTGTAGAACGGGCTGCTCCACTCCCATATGC 420
QY      141 SerPheThrLysGlyGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro 160
Db      421 AGCTTCACCT----- 429
QY      161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
Db      429 ----- 429
QY      181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyrPseProAsn1PheProThr 200
Db      429 ----- 429
QY      201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys 220
Db      429 ----- 429
QY      221 GluIleArgLysGlnGluTyrGlnHisAsnGluValGluTyrAspCysAsnProAsn 240
Db      429 ----- 429
QY      241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrPheThrLeuPro 260
Db      429 ----- 429
QY      261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGlnLeuGlnTyrGlyTyrVal 280
Db      430 -----ATGAAACATGTGATACATACCTGCACTCGAAGTACGATTATGTT 474
QY      281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
Db      475 CAGCCCTCTGCTCCCTCCCTATACACATGAGTTCAGTGTGAGGTGATTTGCGAAATGAA 534
QY      301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleThrPheGlnLeuPro 320
Db      535 TATGCATATGATTTGGAATATACATGATTTACTCTGATTAATGGAATATGACAGAGCTTCT 594
QY      321 MetCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
Db      595 ATG----- 597
QY      341 LeuLeuLysLeuSerGlyLysGlnPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
Db      597 ----- 597
QY      361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGlnValAsp 380
Db      597 ----- 597
QY      381 CysThrGlnLysArgGlnGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
Db      597 ----- 597

```

```

QY      401 MetThrThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
Db      597 ----- 597
QY      421 TyrLeuLeuProGlnAlaLysGlnIleValCysLysAspGlyArgTyrPseLeuPro 440
Db      597 ----- 597
QY      441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnGlnLysAspThr 460
Db      598 ---TGTGTGAGTCTTCTCATATTTGTGGGCCCTCCATCTATTTATACATGAGATACC 654
QY      461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db      655 ACTCATCTCCATTTATACATATATATCTCCAGGGTCAACAGTCAGTACCGTTGCCAGTCC 714
QY      481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrPseGluPro 500
Db      715 TTCTATTAACCTCCAGGGCTGTGTACTGATACATGACAGAAATTAACAGTGTGACAGAACCA 774
QY      501 ProArgCysLeuAspProCysValValSerGlnGluAsnMetAsnLysAsnAsnIleGln 520
Db      775 CCAAGATGCCCTATCATCTGATGTGTATCTGAAGAAACATACACAAATATACATACAG 834
QY      521 LeuLysTyrPheArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGlnPheGln 540
Db      835 TTAATAATGCAAAACAGATGGAACCTATGCAAAAACAGGGGATGCTGTGATATTCGAG 894
QY      541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
Db      895 TGTAAATTCACACATTAACCGATGATATCATCACACCACTTTCGAGCAATCTGTGAGCA 954
QY      561 GlyLysPheGlnTyrProIleCysGlu 569
Db      955 GCGAAATTTGATATCTATATGTGTGA 981

RESULT 5
ABLS7706
ID      ABL57706 standard; DNA; 1095 bp.
XX
AC      ABL57706;
XX
DT      08-OCT-2002 (first entry)
XX
DE      Human sbg614126complfH gene #1.
XX
KW      Human; secreted protein; immunosuppressive; cytostatic; neutrotropic;
KW      neuroprotective; antitumor; vulnerrary; antimicrobial; ophthalmological;
KW      antiparkinsonian; antirheumatic; antiatherosclerotic; dermatological;
KW      hypotensive; cerebroprotective; virucide; antiinflammatory; diabetes;
KW      malignant tumour; hypertension; hypotension; obesity; bulimia; anorexia;
KW      asthma; manic depression; dementia; delirium; mental retardation;
KW      Huntington's disease; Tourette's syndrome; schizophrenia;
KW      mental disorder; sexual development disorder; blood cascade dysfunction;
KW      stroke; growth disorder; gene; ds.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..1095
FT      FT      /tag=a
FT      FT      /product="sbg614126complfH"
XX
PD      WO200222802-A1.
XX
XX      21-MAR-2002.
XX
XX      13-SEP-2001; 2001WO-US28462.
XX      PF
XX      13-SEP-2000; 2000US-232455P.
XX      PR
XX      13-SEP-2000; 2000US-232463P.
XX      PR
XX      02-OCT-2000; 2000US-237293P.
XX      PR

```

PR 07-NOV-2000: 2000US-246269P.  
 PR 20-NOV-2000: 2000US-252049P.  
 XX  
 PA (SMIR ) SMITHKLINE BEECHAM CORP.  
 PA (SMIR ) SMITHKLINE BEECHAM PLC.  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Agarwal P, Gogswell JP, Lai Y, Martensen SA, Smith RF, Strum JC,  
 PI Xie Q;  
 DR WPL: 2002-393963/42.  
 DR P-PSDB: ABB80570.  
 XX  
 PT Novel isolated secreted polypeptides and polynucleotides encoding them  
 PT useful for treating cancer, Alzheimer's disease, tumor metastasis,  
 PT autosomal recessive atypical hemolytic uremic syndrome, wound healing  
 PT disorder  
 XX  
 PS 2; Page166; 246pp; English.  
 XX  
 CC The invention relates to a novel isolated polypeptide (ABB80569-ABB80612)  
 CC (secreted polypeptide) which is encoded by any one of 44 polynucleotide  
 CC sequences (AB157705-AB157748) given in the specification. The  
 CC polypeptides have immunosuppressive, cytostatic, neurotropic,  
 CC neuroprotective, antitumor, vulnerary, antimicrobial, ophthalmological,  
 CC antiParkinsonian, antileukemic, antithrombotic, dermatological,  
 CC hypotensive, cerebroprotective, virucide, and antiinflammatory activity.  
 CC The polynucleotide and polypeptide are useful for treating diabetes,  
 CC malignant tumors, hyper- and hypotension, obesity, bulimia, anorexia,  
 CC asthma, manic depression, dementia, delirium, mental retardation,  
 CC Huntington's disease, Tourette's syndrome, schizophrenia, growth, mental  
 CC or sexual development disorders, and dysfunctions of the blood cascade  
 CC system including those leading to stroke.  
 XX  
 SQ Sequence 1095 BP; 341 A; 217 C; 223 G; 314 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 2.03e-132 Length: 1095  
 Score: 1599.00 Matches: 307  
 Percent Similarity: 56.00% Conservative: 1  
 Best Local Similarity: 55.82% Mismatches: 0  
 Query Match: 50.24% Indels: 242  
 DB: 24 Gaps: 2  
 US-09-805-337A-2 (1-569) x AB157706 (1-1095)  
 QY 20 GlyThrLeuGlyAspPheProLysIleHisGlyPheLeuTyraSpGluIuaSPtyr 39  
 DB 169 GGAACCTTTGGATTTTCCAAAATACACCATGGATTTCTGTATATGAGAGATTTAT 228  
 QY 40 AsnProPheSerGlnValProThrGlyGluValPheTyrtYserCysGluTyraSphe 59  
 DB 229 AACCTTTTCCCAAGTCTCTACAGGGGAAGTTTCTATTACTCCGTGAATATATTTT 288  
 QY 60 ValSerProSerLysSerPheThrPThrArgIleThrCysThrGluGluGlyTrpSerPro 79  
 DB 289 GTGTCCTTCAAAATCCTTTGGACCTGCATTAACATGACAGAAAGAGATGTCACCA 348  
 QY 80 ThrProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 99  
 DB 349 ACACCCAAATGCTTCGAAATGTGTCTCTTCTTTGTGAAAATGGTCATTTCTGATCT 408  
 QY 100 SerGlyLeuIleHisLeuGluGlyAspThrValGlnIleLeuCysAsnThrGlyTrpSer 119  
 DB 409 TCAGGACTATACATCTCGAAGGTGATACGTACAAATATTTCGAACACAGATACACG 468  
 QY 120 LeuGlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProlle 139  
 DB 469 CTTCAAAAACAATGAGAAAACAATTTGTGTAGAAAGGGGGCTGCTCCACTCTCCATA 528  
 QY 140 CysSerPheThrLysGlyGluCysHisValProIleLeuGlnLysAsnValAspAlaGln 159  
 DB 529 TGGAGCTTCACT----- 540

QY 160 ProLysLysGluSerTyrtYrsValGlyAspValLeuLysPheSerCysArgLysAsnLeu 179  
 DB 540 ----- 540  
 QY 180 IleArgValGlySerAspSerValGlnCysTyrglnPheGlyTrpSerProAsnPhePro 199  
 DB 540 ----- 540  
 QY 200 ThrCysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyGlnVal 219  
 DB 540 ----- 540  
 QY 220 LysGluIleArgLysGluGluTyrglyHisAsnGluValValGlyTyraSPcysAsnPro 239  
 DB 540 ----- 540  
 QY 240 AsnPheIleTLeasnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeu 259  
 DB 540 ----- 540  
 QY 260 ProThrCysValGluGlnValLysThrCysGlyTyrtIleProGluLeuGluTyrglyTyrt 279  
 DB 541 -----ATGAAACATGCTGATACATACCTGACCTCAGTACGCTTAT 582  
 QY 280 ValGlnProSerValProProTyrglnHisGlyValSerValGluValAsnCysArgAsn 299  
 DB 583 GTTCACCCCTGCTGCTCCCTCCATACATGAGATTTGCTGAGCTGATATTCAGCAAT 642  
 QY 300 GluTyraIleMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeu 319  
 DB 643 GAATATGCAATATTTGGAATTAACATGATTAACCTGATTAATGAATATGAGACAGCTT 702  
 QY 320 PrometCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLys 339  
 DB 703 CCTATG----- 708  
 QY 340 ThrLeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyraGlyCys 359  
 DB 708 ----- 708  
 QY 360 SerAspIlePheArgTyraGlnHisSerValCysIleAsnGlyLysTrpAsnProGluVal 379  
 DB 708 ----- 708  
 QY 380 AspCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGln 399  
 DB 708 ----- 708  
 QY 400 AsnMetThrThrValAsnTyrglnAspGlyGluLysValAlaValLeuCysLysGlu 419  
 DB 708 ----- 708  
 QY 420 AsnTyrlLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeu 439  
 DB 708 ----- 708  
 QY 440 ProArgCysValGlnSerThrAlaTyrcysGlyProProProSerIleAsnAsnGlyAsp 459  
 DB 709 -----TGTTGTAGCTTACTGTCATATTGTGGGCCCTCCATCTATTACAAATGAGAT 762  
 QY 460 ThrThrSerPheProLeuSerValTyrtProProGlySerThrValThrTyraGlyCysGln 479  
 DB 763 ACACCTCATTTCCCATATATCATATATCTCCAGGGTCAACAGTACGTTGCCAG 822  
 QY 480 SerPheTyrlLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGlu 499  
 DB 823 TCCCTCTATAAATCCAGGGCTCTGTAACTGTAACATGAGAAATTAACAGTGGTCAGAA 882  
 QY 500 ProProArgCysLeuAspProCysValValSerGluGlnLysAsnMetAsnLysAsnIle 519  
 DB 883 CCACCAAGATGCTATGATCATATGTGTGTATGAGAAACAATGAAACAATAATACATA 942





Db 948 AAGAAGATGATACACACAGCTCTGCATTAATGGAAGATGGGATCCAGAGTGAATGCTC 1007  
 QY 382 ThrGluysArgGluGlnPheCysProProProGlnIleProasnaIaGlnasmet 401  
 Db 1008 TCAATGGCCCAATATACATTTATGCGCACCCTCCACCTCCAGATTCCCATTTCTCACAATATG 1067  
 QY 402 ThrThrValAsnTyrGlnAspGlyGluysValAlaValLeuGlySylsGluAsnTyr 421  
 Db 1068 ACAACCACTGAATATTCGGATGGAGAAAAAGTATCTGTTCTTGGCAAGAAATATAT 1127  
 QY 422 LeuLeuProGluAlaIaIysGluIleValGlySylsAspGlyArgTyrGlnSerLeuProArg 441  
 Db 1128 CTAAATTCAGGAAGAGAGAAATATACATGCAAAAGATGAGATGGCATATACACATC 1187  
 QY 442 CysValGlu----- 444  
 Db 1188 TGTGTGAAAAAATTCACATGTTCCACACACCTCAGATAGAACAGGAACATTATATCA 1247  
 QY 444 ----- 444  
 Db 1248 TCCAGGCTTTCACAGAAGATTATGCACATGGAGCTAATATGATTATCTGTGAGGGT 1307  
 QY 444 ----- 444  
 Db 1308 GGTTCAGAGATATCTGAGAGAAATGAAACACATGCTACATGGGAAATGAGCTTCTCA 1367  
 QY 444 ----- 444  
 Db 1368 CCTCACTGCAAGCGCTTCTCTTGTAAATCTCCACCTGACATTTCTCATGGTGTAGCT 1427  
 QY 444 ----- 444  
 Db 1428 CACATGTCAGACAGTATCATGATGAGAGAAAGATTACAAATGTTTGAAGSTTTT 1487  
 QY 444 ----- 444  
 Db 1488 GGAATTGATGGCGCTGCATTCGCAAAATGCTTAGAGAAAAATGTCACCTCCATCA 1547  
 QY 444 ----- 444  
 Db 1548 TGCATAAAAAGATTGTCTCAGTTTACCTAGCTTGAATGGCATACCATGGCAGAG 1607  
 QY 444 ----- 444  
 Db 1608 AAGAAGATGTATATAGCGGGGTGAGCAAGTGCATTACCTTGCAACATATTACAAA 1667  
 QY 444 ----- 444  
 Db 1668 ATGGATGAGCCAGTAAATGTAACATCATTAATAGCAGATGACAGGAAGCCCAACATGC 1727  
 QY 444 ----- 444  
 Db 1728 AGAGACACTCTCTGTGTGATCCGCCACAGTACAAATTCCTTATATAGTGTGAGACAG 1787  
 QY 444 ----- 444  
 Db 1788 ATGAGTAATATTCATCTGTGTGAGAGATAGCTTATCATGTAGAGCCCTTATGAATG 1847  
 QY 444 ----- 444  
 Db 1848 TTGGGGATGAAGAGTATGTGTTTAAATGGAAGTGAACGACACCTCATSCAAA 1907  
 QY 445 ---SerThrAlaTyrCysGlyProProSerIleAsnAsnGlyAspThrThrSerPhe 463  
 Db 1908 GATTCACAGAAATGTGGGCCCTCCACCTATTGACAAATGGGACATTAATTCATTC 1967  
 QY 464 ProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSerPheTyrLys 483  
 Db 1968 CCGTTGTAGTATATCTCCAGCTCATGATGATGACCAATGACAGAACTGTATCA 2027  
 QY 484 LeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrSerLysProProArgCys 503  
 Db 2028 CTTAGGGGTAAACAGCAATATACATGTGAAATGACAAATGTCACAGACCAACCAAAATGC 2087

QY 504 LeuAspProCysValIaIserGluIaAsnMetAsnLysAsnIleGlnLeuLysTrp 523  
 Db 2088 TTATGATCCCTGTGTATATCCGAGAGAAATTTATGAGAAATTTATACATTAAGTGG 2147  
 QY 524 ArgAsnAspGlyLysLeuTyrAlaIysThrGlyAspAlaValGluPheGlnCysLysPhe 543  
 Db 2148 ACAGCCAAACAGACCTTATTCGAGAACAGGTGATCAGTGAATTTGTGTAAACGG 2207  
 QY 544 ProIleLysAlaMetIleSerSerProProPheArgAlaIleCysGlnIleLysPhe 563  
 Db 2208 GGAATATGCTTTTCATCACTTCATCCACATTCGAAACAACATGTTGGAGTGGAAACGG 2267  
 QY 564 GluTyrProIleCys 568  
 Db 2268 GAGTATCCAACTTGT 2282  
 RESULT 7  
 AAC77947 standard; cDNA: 1293 BP.  
 AAC77947:  
 AC  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human cancer associated gene sequence SEQ ID NO:341.  
 KW Human; cancer associated gene; cancer antigen; detection; cancer;  
 KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;  
 KW antidiabetic; antihistaminic; antihemetic; antichratic; antiviral;  
 KW antineoplastic; antihypertensive; antiallergic; antibacterial; anti-inflammatory;  
 KW dermatological; neuroprotective; thrombolytic; coagulant; neotropic;  
 KW vasotropic; antiproliferative; angiogenic; gene therapy; inflammation;  
 KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
 KW allergic reaction; graft versus host disease; organ rejection;  
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
 KW neurological disease; drug screening; ss.  
 OS Homo sapiens.  
 XX  
 OS  
 PN MO20005350-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000MO-US05882.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 PI Rosen CA, Ruben SM;  
 PI  
 DR WPI: 2000-587533/55.  
 DR P-PSDB: AAB43738.  
 XX  
 PT Novel isolated nucleic acids comprising sequences encoding peptides  
 PT useful for treating or diagnosing e.g. cancer -  
 XX  
 PS Claim 1; Page 892; 2352pp; English.  
 XX  
 CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
 CC in AAB43398 to AAB44239. The proteins can have activities based on the  
 CC tissues and cells the genes are expressed in. Example of activities  
 CC include: cytostatic; proliferative; vulnary; immunomodulator;  
 CC antidiabetic; antihistaminic; antihemetic; antichratic; antiviral;  
 CC anti-inflammatory; antihypertensive; antiallergic; antibacterial; anti-inflammatory;  
 CC dermatological; neuroprotective; cardiatic; thrombolytic; coagulant;  
 CC neotropic; vasotropic; antiproliferative; angiogenic; gene therapy; inflammation;  
 CC polynucleotides and polypeptides can be used for preventing, treating or  
 CC ameliorating medical conditions and diagnosing pathological conditions.  
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
 CC the present invention may be used to treat immune disorders by activating  
 CC or inhibiting the proliferation, differentiation or mobilisation of

CC Immune cells, to treat disorders of hematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, CC agonists and antagonists may be also be used in drug screens. AAC78449 to CC AAC78437 and AAB44240 represent sequences used in the exemplification of CC the present invention.

XX Sequence 1293 BP; 427 A; 242 C; 247 G; 367 T; 10 other:

Alignment Scores:

Pred. No.:	1,07e-83	Length:	1293
Score:	1051.50	Matches:	216
Percent Similarity:	43.31%	Conservative:	30
Best Local Similarity:	38.03%	Mismatches:	81
Query Match:	33.03%	Indels:	241
DB:	21	Gaps:	5

US-09-805-337A-2 (1-569) x AAC7947 (1-1293)

```

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 89 ATGTGGCTCTGTGTGAGTAACTTAATCTCAGCATATCTCTGTGGGAGAGACA 148
QY 21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyraSpGluLysPtyrAsn 40
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 149 ACATTTTGTGATTTTCCAAAATAAACCATGCAATCTATATGATGAAGAAAATATAG 208
QY 41 ProPheSerGluValProThrGlyGluValPheTyTrSerCysGluTyraSpPheVal 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 209 CCATTTTCCAGAGTCTCTCAGAGGAGATTTTCTATCTCTCTGTAATATATATTTTG 268
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluLysTrpSerProThr 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 TCTCCTTCAAAATCATTTTGGACTCGCATACATGCAGAGAAGAGATGTCACCAACA 328
QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 329 CCAAAATGTCAGACTGTGTCTTCTTGTGGAATAAGTCATCTCAATCTTCA 388
QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 389 GGACAAACACATCTGGAAGGTGATCTGTGCAAAATTAATTTGCACACAGATACAGACTT 448
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProIleCys 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 449 CAAACAAATGAGAACAACTTTCATCTGTAGAACGGGCTGTCCACCCCTCCAAATGC 508
QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 509 AGGTCCACTGAC----- 520
QY 161 LysLysGluSerTyrlsValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 520 ----- 520
QY 181 ArgValGlySerAspSerValGlnCysTyrlGlnPheGlyTrpSerProAsnPheProThr 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 520 ----- 520
QY 201 CysLysGlyGlnValArgSerCysGlyProProProGlnLeuSerAsnGlyLysValLys 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 521 -----ACTTCCTGTGTGAATCCGCCACACATGACAAATGCTATATATATAG 565
QY 221 GlnIleArgLysGluLysArgLysHisAsnGluValGlnIleTyraSpCysAsnProAsn 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 566 TCGACAGACATGATGTAATATCATCTGTGTGAGACAGTACGTTAT----- 610
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 610 ----- 610

```

```

QY 261 ThrCysValGluGlnValLysThrCysGlyTyrlIleProGluLeuGluTyrlVal 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 610 ----- 610
QY 281 GlnProSerValProProTyrlGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 -----SAATGTAGAGGAGCCCT 625
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 626 TATGAATGTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 679
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 679 ----- 679
QY 341 LeuLeuLysLeuSerGlyLysGluLysPheAsnHisAsnSerArgIleArgTyrlArgCysSer 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 679 ----- 679
QY 361 AspIlePheArgTyrlArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 679 ----- 679
QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 680 -----CCACCTCAA----- 688
QY 401 MetThrThrThrValAsnTyrlGlnAspGlyGluLysValAlaValLeuCysLysGluAsn 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 688 ----- 688
QY 421 TyrlLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 689 -----TCAAAAGAT----- 697
QY 441 ArgCysValGluSerThrAlaTyrlCysGlyProProProSerIleAsnAsnGlyAspThr 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 698 -----TCTACRGAATAATGTGGGCCCTCCACCTATGACAAATGGGACATT 745
QY 461 ThrSerPheProLeuSerValTyrlProProGlySerThrValThrTyrlArgCysGlnSer 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 746 ACTTCATTCCTGTGTGATATATCTGTGATATATCTGTGATATATCTGTGATATATCTGTGAT 805
QY 481 PheTyrlLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 806 TGTATCAACTTGAGAGGTATACAGCAATATACATGTAGAAATGAGCAATGTGCAGACACA 865
QY 501 ProArgCysLeuAspProCysValLysSerGluGluAsnMetAsnLysAsnIleGln 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 866 CCAAAATGCTTACATCCCTGTCTATATATCCCGAATAATTAATTAACATACCA 925
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTyrlAlaLysThrGlyAspAlaValGlnPheGln 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 926 TTAAGGTGACAGCCCAACACAGACGTTATTTTGAAACAGCGAATCACTGATCAATTTTGTG 985
QY 541 CysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGlnGlu 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 986 TGTAAACGGGATATATGCTTTCATACAGTCTGCACATTCGCAACAAACATGTCGGAT 1045
QY 561 GlyLysPheGluTyrlProIleCys 568
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1046 GGGAACTGAGATATCAACTTGT 1069

```

RESULT 8

ABN95888

ID ABN95888

standard. DNA: 1266 BP.

ABN95888;

13-AUG-2002 (first entry)

DE Gene #2386 used to diagnose liver cancer.





```

Db      780 AGAGTCGAGTACGAGTCCTACTGAACTTCAGGTCCTCAATATGTAACATGT 839
QY      252 ValAspGlyGluThrPheProthrycysValGluGlnValLysThrCysGlyTyr 271
Db      840 ACTAATGGAGACTGCGACCAACCAAGATGCATA--TCAATGAACCTTGAGTTT 896
QY      272 IleProGluLeuGluTyrGly-----TyrValGlnProSerValProProTyr----- 287
Db      897 ---CCAGAAATTCACATGACATCTATTTATTTAGAAATACCGCTGACCATTTCCA 953
QY      288 ---GlnHisGlyValSerValGluValAsnCysArgAsnGluTyrAlaMetIleGlyAsn 306
Db      954 GTAGCTACAGCAACTTACTCTCTATTTACTGTGACAAATTTGTGACTCTCCAGCA 1013
QY      307 -----AsnMetIleThrCysIleAsnGlyIleTyrPheGluLeuProMetCysVal 323
Db      1014 ACTTACTGGATTACTTCACTTCACGACACAGATGGGTGGTCCAAAGCTCCCATC--- 1070
QY      324 AlaThrHisGlnLeuLysArgCysIleAlaGlyValAsnIleLysThr--LeuLeu 342
Db      1071 -----CTCAGAACATGCTCCAAATGACATATGAAATTTGAAATGATTCATT 1118
QY      343 LysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSerAspIle 362
Db      1119 TCTGAATCTCTCTATTTATTTATTTAAATAAAGAAATATCAATATGTAACCAAGA 1178
QY      363 PheArgTyrArgHisSer-----ValCysIleAsnGlyLysTyrPasn 376
Db      1179 TATGACACGACAGATGAAATTCCTGAGTTCAATTACATGATTTGGAAATGATGATCA 1238
QY      377 ProGluValAspCysThrGluLysArgGluPheCysProProProGluIlePro 396
Db      1239 GCACACCAATTTGCATT-----AAATTTGT---GATATGCTGTTTGTGAG 1283
QY      397 AsnAlaGlnAsnMetThrThrValAsnTyrGlnAspGlyGluLysValAlaValIleu 416
Db      1284 AATTCAGAGCCAGAGATAGGATGCGATGTTTAACTCCATGACATGCGACTGCGAA 1343
QY      417 CysLysGluAsnTyrLeuLeuProGluAlaLys-----GluIleValCysLysAsp 433
Db      1344 TGTCTACGATGATATATAATTCAGTTTGGAAACACCAAGCTCCATAGTGTGTGTGA 1403
QY      434 GlyArgTyrPheLysLeuPheProArgCysValGluSerThrAlaTyrCysGlyProPro 453
Db      1404 GATGGGTGTCCTCCATTTCCACACATGTTPMAATTTTCAGAAAGTGGGCTCCTCCA 1463
QY      454 SerIleAsnAsnGlyAspThrThrSerPheProLeuSerValTyrProProGlySerThr 473
Db      1464 CCTATTAGCAATGGTATACACACCTCTTCTACTAAAGTGTATGTGCCACGATCAGA 1523
QY      474 ValThrTyrArgCysGlnSerPheTyrLysLeuGlnGlySerAlaThrValThrCysArg 493
Db      1524 GTGCGATACCAATGCCAGTCTTACTATGAACTTCAGGTTCTATATATGATACATGTAGT 1583
QY      494 AsnLysGlnTyrSerGluProProArgCysLeuAspProCysValValSerGluGluAsn 513
Db      1584 AATGAGAGGTGGTGGAGACCAACCAAGATCATGTATATTAATCTGAAGAAAG 1643
QY      514 MetAsnLysAsnAsnIleGlnLeuLysTyrPargAsnAspGlyLysLeuTyrAlaLysThr 533
Db      1644 ATGAAATAAATAACATACAGTTTAAAGAAAGAAAGACATTAATATATGCAAAAACCA 1703
QY      534 GlyAspAlaValGluPheGlnCysLysPheProHisLysAlaMetIleSerSerProPro 553
Db      1704 GGGGATACCAATTAATTTATGTATTAATTTGGGATATAAATGCGATACATGATTCATCA 1763
QY      554 PheArgAlaIleCysGlnGluGlyLysPheGluTyrProIleCysGlu 569
Db      1764 TTTCAGACGTGTGTAGGAGGCAATAGTGGAATACCCACAGATGCGAA 1811

```

RESULT 10  
ABN97268

```

ID      ABN97268 standard; DNA; 1040 BP.
XX
AC      ABN97268;
XX
DT      13-AUG-2002 (first entry)
XX
DE      Gene #3766 used to diagnose liver cancer.
XX
KW      Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW      metastatic liver tumor; cytostatic; expression profile; disease state;
KW      disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS      Homo sapiens.
XX
PN      W0200229103-A2.
XX
PD      11-APR-2002.
XX
PF      02-OCT-2001; 2001WO-US30589.
XX
PR      02-OCT-2000; 2000US-237054P.
XX
PA      (GENE-) GENE LOGIC INC.
XX
PI      Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX      WPT; 2002-426119/45.
XX
PT      Diagnosing and detecting the progression of liver cancer,
PT      hepatocellular carcinoma or metastatic liver tumor in a
PT      liver tissue sample -
XX
PS      Claim 1; SEQ ID NO 3766; 298bp; English.
XX
CC      The invention relates to a novel method for diagnosing and detecting the
CC      progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC      tumor in a patient, and differentiating metastatic liver cancer from
CC      hepatocellular carcinoma in a patient, involving detecting the level of
CC      expression of two or more genes represented in ABN93503-ABN97455 in a
CC      tissue sample. The method of the invention has hepatotropic, and
CC      cytostatic activity. The method is useful for diagnosing and detecting
CC      the progression of liver cancer, hepatocellular carcinoma and metastatic
CC      liver carcinoma in a patient. The method is useful for identifying
CC      expression profiles which serve as useful diagnostic markers as well as
CC      markers that can be used to monitor disease states, disease progression,
CC      drug toxicity, drug efficacy and drug metabolism.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
SO      Sequence 1040 BP; 348 A; 192 C; 185 G; 315 T; 0 other;
XX
Alignment Scores:
Pred. No.:      8.07e-73      Length:      1040
Score:          927.50      Matches:      198
Percent Similarity: 38.1%      Conservative: 20
Best Local Similarity: 34.62%      Mismatches: 47
Query Match:    29.14%      Indels:      307
DB:            24      Gaps:      3
US-09-805-337A-2 (1-569) x ABN97268 (1-1040)
QY      1 MetLeuLeuLeuPheSerValIleLeuIleSerTyrValSerThrValGlyCysGly 20
Db      78 AGTGGCTCCTGCTCGATTAATTCATCTCAGGATATCTCTTGCGGAGCAAGA 137
QY      21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsn 40
Db      138 ATGTTCTGATTTTCCAAAATAAACCATGGAATTCATATGATATAAGAAATATTAAG 197
QY      41 ProPheSerGlnValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60

```

```

Db 198 CCAATTTTCCCAAGTTCTACAGCGAAGTTTCTATTACTCTCGTAATATATTTTGTC 257
QY 61 SerProSerLysSerPheThrPheThrCysThrGluGluGlyTrpSerProThr 80
Db 258 TTCTCTTCAAAATCTCTTTGGACTCCCATACGTGCGAGAAAGAGATGTCACCAACA 317
QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100
Db 318 CCAAGTGTCTCAGACTGTCTTCTTCTTCTGCGAAATGCGCATCTGATCTGCA 377
QY 101 GlyLeuLeuHisLeuGluGluGlyAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120
Db 378 GGACAACAACATCTCGAAGGTATCTACTATCAAAATATTTGCCACACAGATACAGACT 437
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys 140
Db 438 CAAACCAATGACACACACACTTTCATGTCATGACACGGGGTGGTCCACTCTCCCAATGC 497
QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160
Db 497 ----- 497
QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
Db 497 ----- 497
QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProThr 200
Db 497 ----- 497
QY 201 CysLysGlyGluValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys 220
Db 497 ----- 497
QY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240
Db 497 ----- 497
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrLeuPro 260
Db 497 ----- 497
QY 261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyTyrVal 280
Db 497 ----- 497
QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnGlyAsnGlu 300
Db 497 ----- 497
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320
Db 497 ----- 497
QY 321 MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyAlaAsnIleLysThr 340
Db 497 ----- 497
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
Db 497 ----- 497
QY 361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380
Db 497 ----- 497
QY 381 CysThrGluLysArgGluGlnPheCysProProProProGlnIleProAsnAlaGlnAsn 400
Db 497 ----- 497
QY 401 MetThrThrValAsnTyrGlnAspGlyGluLysValAlaValLeuLysCysLysGluAsn 420
Db 497 ----- 497

```

```

QY 421 TyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro 440
Db 497 ----- 497
QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAspThr 460
Db 498 AGTCCACCATTTCTGCAGAAAAATGTGGGCCCTCCACCATATTCAGCAATGAGACATT 557
QY 461 ThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGlnSer 480
Db 558 ACTTATTCCTGCTGTTCAGTATATATCTCCAGGTCATCATGTAAGTACCGACGACGAGAC 617
QY 481 PheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTrpSerGluPro 500
Db 618 TTGTATCAACTGACAGGTATACAAATTAACATGATAGAAAGCAGCAATGTGCACAAACA 677
QY 501 ProArgCysLeuAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIleGln 520
Db 678 CCAAAATGCTTAGATCCATGCTATATATCCACAGAAATTAATGCAAAATATACATAAAA 737
QY 521 LeuLysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPheGln 540
Db 738 TTAAGTGCACAAACCAACAAAGCTTTATTCAGAACAGGTGACATGTTGAATTTGTT 797
QY 541 CysLys-----PheProHisLysAlaMetIleSerSerProProPheArgAlaIle 557
Db 798 TCTAATCTGATATCATCCACAAACAA-----TTCATTCATTTCCGACCAATG 845
QY 558 CysGlnGluGlyLysPheGluTyrProIleCysGlu 569
Db 846 TGTCAAAATGGGAACGTGTATATCCAGTTGTGAG 881

RESULT 11
AAN82402
ID AAN82402 standard; DNA; 2177 BP.
XX
AC AAN82402;
XX
DT 26-NOV-1990 (first entry)
XX
DE B subunit of human factor XIII.
XX
KW B sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
KW crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
FH key Location/Qualifiers
FT CDS 2..1521
FT FT /*tag= a
FT FT /product= b subunit
FT FT misc_RNA 2..58
FT FT /*tag= b
FT FT /label= leader_sequence
FT FT CDS 59..1521
FT FT /*tag= c
FT FT /product= mature b subunit
PN AUB778694-A.
XX
PD 31-MAR-1988.
XX
PE 21-SEP-1987; 87AU-0078694.
XX
PR 19-SEP-1986; 86US-090512.
XX
PA (ZYMO-) ZYMOGENETICA INC.
XX
PI Davie E, Seale RL, Ichinose A, Holly JA, Parker GE;
XX WPI: 1988-140637/21.
XX DR
DR P-PSDB: AAB82921.

```

XX New DNA sequences encoding sub-units of factor 13 - and corresponding  
PT expression vectors and transformed host cells.  
XX  
PS Disclosure: ; P; English.  
CC The carboxyl-terminal Thr (nucleotides 1979-1981) is followed by  
CC a stop codon (TAG), a 187 bp noncoding sequence, and a poly(A) tail  
CC of 9 bp. The polyadenylation or processing signal of AATAAA  
CC was identified 19 nucleotides upstream from the poly(A) tail.  
CC When cultured the host cells will produce the polypeptide which  
CC can be assembled to factor 13, which stabilises blood clots by  
CC crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.  
CC in cases of scleroderma, haemorrhage, ulcerative colitis etc., and  
CC can be prepod. in large aunts. Without risk of viral contamination.  
CC See also AAN82401 and AAN82403.  
XX  
SQ Sequence 2177 BP; 758 A; 358 C; 434 G; 627 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 3,48e-57 Length: 2177  
Score: 756.50 Matches: 194  
Percent Similarity: 41.46% Conservative: 85  
Best Local Similarity: 28.83% Mismatches: 235  
Query Match: 23.77% Indels: 139  
DB: 9 Gaps: 21  
  
US-09-805-337A-2 (1-569) x AAN82402 (1-2177)  
OY 3 LeuLeuPheSerValIleLeuIleSerTTPvalSerThValGlyGlyLeu 22  
DB 14 CTGACTTTATCATCATATTGATA-----ATCTCAGAGAA----- 49  
OY 23 CysAspPheProLysIleHisSGLyPheLeuTyfAspGluLysPyrAsnProPhe 42  
DB 50 -----CTCTATGCAAGAGAGAAACCCCTGTGCTTT 79  
OY 43 SerGlnValProThrGlyGlyVal-----PheTyfTyrSer----- 54  
DB 80 CCTCATGTGGAATAATGAGAAATTGGCCCAATATTACATATTAAACCTTTACTTT 139  
OY 55 -----CysGluTyfAsnPheValSerProSer 63  
DB 140 CCAATGAGCATACACAAAAATTTGCTATTTTCTGCTGGCTGTTATACCACTGAAGT 199  
OY 64 LysSerPheThrPheArgIleThrCysThrGluGlyIleTyrSerProThrProLysCys 83  
DB 200 GGAAGACAAAGAGAGCAAAACGCTGTACACAGAAAGGCTGCTCTCCAGGCCAAGTGC 259  
OY 84 LeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSerGlyLeuIle 103  
DB 260 TTCAAATAATGCACTAAGCCTGACCTGAGTAATGCTTACATCTCTGATGTAAGTTATG 319  
OY 104 His----- 104  
DB 320 TATTAATTCAGAGAACATGATATGTTGGCTTCAGGGTACAAACCACTGAGAGG 379  
OY 105 -----LeuGluGly----- 107  
DB 380 AAGGATGAAGAAGTGGTTCAATGCTCTCTGATGATGCTCTTCTACACCAACCTGAGG 439  
OY 107 ----- 107  
DB 440 AAGAACATGAAMCGTTTGGCTCGATTTATATATGAATTTATTCACACACAG 439  
OY 108 -----AspThrValGlnIleIleCysAsnThrGlyTyrSerLeuGln 121  
DB 500 AAACATTCAAAGTGAAGCAAGTACATACGAATGTCCTACTGCTGCTACACAGCT 559  
OY 122 AsnAsnGluLys-----AsnIleSerCysValGlnArgGlyThrSerThrProPhe 139  
DB 560 GGAAGAAAGACAGAGAGGATGATGCTCACTATGAGATGCTCTCACACCAAAA 619

OY 140 CysSerPheThrLysGlyLysCysHisValProIleLeuGlnAlaAsnValAspAlaGln 159  
DB 620 TGT-----ACCAATTAAGTGTCTCTTAAAGTATATTAATGAATGTTATTTTCAT 673  
OY 160 ProLysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeu 179  
DB 674 CTGTAAACCAAAACCTATAGAGAGATGCTTCAGTTTCTGTCATGAATAATATAT 733  
OY 180 IleArgValGlySerSerPheValGlnCysTyfGlnPheGlyTyrSerProAsnPro 199  
DB 734 TATCTAAGTGATCTGATTTATTAATCATCTTAATCTTGTGGTACCAAGATCTCT 793  
OY 200 ThrCysLysGlyGlnValAlaArgSerCysGlyProProGluLeuSerAsnGlyVal 219  
DB 794 GTATGCGAAGAGAAAGAAACAGATCTCTCCCTCCACCTCTGCCATTAATCCAAAT 853  
OY 220 LysGluIleArgLysGluGlyTyfGlyHisAsnGlnValGlyTyfAspCysAsnPro 239  
DB 854 ---CAAAACATTCACAACTTATCTCATGAGAAATAGTTCATATGAAATCTGAACCT 910  
OY 240 AsnPheIleLeuAsnGlyProLysIleGlnCysValAspGlyGluThrThrLeu 259  
DB 911 AATTGTGACATCCATGGCTGACGACAAATACGTTGGAAGGTGGAATAAGACACCT 970  
OY 260 ProThrCysValGlu-----GlnValLysThrCysGlyTyfIleProGluLeuGluTyf 277  
DB 971 CCAAAATGCATTTAGACGACAGAGAAAGTACCTCTGAGAGAACCCATTCATGTAAT 1030  
OY 278 GlyTyfValGlnProSerValProProTyfGlnHisGlyValSerValGlnValAsn 297  
DB 1031 GGTGACAGCAATTTTACACTTAAGATTATTAAATGAGGATTAATGACATATGATGT 1090  
OY 298 ArgAsnGlyTyfValMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleThr 317  
DB 1091 AAAGCGGCTACTCTTCTCCATGATGATCCATGAATATCTTGTATATCTGGAATGACA 1150  
OY 318 GluLeuProMetCysValAlaThrHisGlnLeuLysArgCysLys----- 332  
DB 1151 CTCTCTCTGAGAGTGTGTGAATAATATAG-----AATGTATAGACCTCTCTGTTGTA 1204  
OY 333 IleAlaGlyValAlaAsnIleTyfThrLeuLeuLysLeuSerGlyGluPheAsnHisAsn 352  
DB 1205 ATGAATGGGCTGTTCAGACGGGATATTG-----GCAAGCTATGCAACAGCA 1252  
OY 353 SerArgIleArgTyfArgCysSerAspIlePheArgTyfArgHisSerVal----- 369  
DB 1253 TCCTCAGTGAATATATGATGATGATATATCTTACTGAGGAGGATCAAAAATATCTCGT 1312  
OY 370 CysIleAsnGlyLysThrPheProGluValAspCysThrGluLysArgGluGlnPheCys 389  
DB 1313 TCGGACAAAGAAATGCGATCCCACTGTTGCTGGAACCA-----TGT 1360  
OY 390 ProProProGluIleProAsnAlaGlnAsnMetThrThrTyfValAsnTyfGln----- 408  
DB 1361 ACTGTATATGTGATTAACATGACAGAAATATGATGAAATGGAATATATGGAAGG 1420  
OY 409 -----AspGlyGluLysValAlaValLeuCysLysGluSerTyr----- 421  
DB 1421 AAGTCTTACATGAGATTAATATAGATTTTGTATGTAACAGGAGATATGACTTATCCCA 1480  
OY 422 LeuLeuProGluLys-----GluIleValCysLysAspTyfArgThrPheLeuPro 440  
DB 1481 TTAACCCCATTTGCTGATATATCTTGAGCTGACAGAGAGACA-----GTGAATATCC 1537  
OY 441 ArgCysVal-----GluSerThrAlaTyfCysGlyProProProSerIleAsnAsnGly 458  
DB 1538 TTATGTACTAGAAAGAAATCTTAAGAGATGCGACATCTCTCTTATTAAACATGGA 1597  
OY 459 AspThrThrSerPheProLeuSerValTyfProProGluSerThrValThrTyfArgCys 478  
DB 1598 GTCATTTATTTAGTCAACAGTACGACCTTATGAAATGCTCTCAGTATGATATACATGT 1657  
OY 479 GlnSerPheTyfLysLeuGlnGlySerValThrValThrCysArgAsnLysGlnTyrSer 498





```

Db 582 -----TCCTCAGCATCAATTACATGTTCCGA----- 608
OY 320 PrometCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLys 339
Db 608 ----- 608
OY 340 ThrLeuLeuLysLeuSerGlyLysGlnPheAsnHisAsnSerArgIleArgTyrArgCys 359
Db 608 ----- 608
OY 360 SerAspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluVal 379
Db 609 -----AATGGA----- 614
OY 380 AspCysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGln 399
Db 614 ----- 614
OY 400 AsnMetThrThrValAsnTyrGlnAspGlyLysValAlaValLeuCysIysGln 419
Db 614 ----- 614
OY 420 AsnTyrLeuLeuProGluAlaLysGluIleValCysLysAspGlyArgTyrGlnSerLeu 439
Db 615 -----TGGTCAGCACCA 626
OY 440 ProArgCysValGluSerThrAlaTyrCysGlyProProProSerIleAsnAsnGlyAsp 459
Db 627 CCAATTGGATTAAATTTCTTCAGAAAAGTGTGACCTCTCCACCTATTGACANTGGTGTAT 686
OY 460 ThrThrSerPheProLeuSerValTyrProProGlySerThrValThrTyrArgCysGln 479
Db 687 ACCACCTCCTTCTCTACTAAAGTATGATGCGCACAGTCAAGAGTGCAGTACCAATGCCAG 746
OY 480 SerPheTyrLysLeuGlnGlySerValThrValThrCysArgAsnLysIleThrSerGlu 499
Db 747 TCCCTACTAGAACCTTCAGGGTTCCTAATTATGTGATACATGATGATGAGAGGTCGGCA 806
OY 500 ProProArgCysLeuAsnAspProCysValValSerGluGluAsnMetAsnLysAsnAsnIle 519
Db 807 CCACCTAGATGATCATCATCATGATATTAATACTGAAAGAAACATGATATAAATAATACATA 866
OY 520 GlnLeuLysTyrArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaValGluPhe 539
Db 867 AAGTTAAAGAAAGAAAGTATGATATTCGCAAAAACAGGGGATACCATTTGAATT 926
OY 540 GlnCysLysPheProHisLysAlaMetIleSerSerProProPheArgAlaIleCysGln 559
Db 927 ATGTGTAAATTGGATATATGCAAAATACATCAATTCTATTCATTCACAGAGTGTGCG 986
OY 560 GluGlyLysPheGluTyrProIleCysGln 569
Db 987 GAAGGATAGTGATATCCCGCATGCCGA 1016

```

RESULT 13  
AAS06029 standard; DNA; 1929 BP.

12-SEP-2001 (first entry)

Angiotensin converting enzyme (ACEV) splice variant DNA #29.

Angiotensin converting enzyme splice variant; ACEV: interleukin 6; granulocyte colony stimulating factor receptor; glucagon; hypertrophy; platelet-derived endothelial cell growth factor; cardiovascular disease; cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1c; ds; vasorelaxant intestinal polypeptide receptor 2; arteriosclerosis; cancer; myocardial infarction; coronary arterial thrombosis; renal disease; diabetic nephropathy; muscular disease; immune disorder; sarcoidosis; multiple sclerosis; immune complex nephritis; deep vein thrombosis; nonatroidic pulmonary granulomatous disease; endothelial abnormality;

vascular disorder; asbestosis.

Mus sp..

MO200136632-A2.

25-MAY-2001.

17-NOV-2000: 2000MO-1100766.

17-NOV-1999: 991L-0132978.

10-DEC-1999: 991L-0133455.

(COMP-) COMPUGEN LTD.

Levine Z, David A, Azar I, Khosravi R, Bernstein J;

WPI; 2001-336004/35.

P-PSDB; AAU02929.

Novel alternative splicing variants e.g. variant of angiotensin converting enzyme (ACEV) useful in identifying candidate compounds capable of binding to the variant and to detect anti-variant antibodies

Claim 1: Page 330-331; 519pp; English.

The sequence represents a DNA encoding an angiotensin converting enzyme splice variant (ACEV) polypeptide. The polypeptides of the invention include variants of granulocyte colony stimulating factor receptor, glucagon, interleukin 6, platelet-derived endothelial cell growth factor, cyclin-dependent kinase inhibitor 1c, cellular tumour antigen p53, and vasorelaxant intestinal polypeptide receptor 2. The polypeptides and their associated nucleic acids are useful for identification of variant sequences and detection of candidate compounds capable of binding the molecules. The sequences of the invention can be used in the treatment and diagnosis of various disorders including cardiovascular diseases such as arteriosclerosis, myocardial infarction and coronary arterial thrombosis, renal diseases such as diabetic nephropathy, muscular diseases such as hypertrophy, immune disorders such as immune complex nephritis, multiple sclerosis, cancer, sarcoidosis, nonatroidic pulmonary granulomatous diseases such as asbestosis and vascular diseases involving an endothelial abnormality such as deep vein thrombosis.

Sequence 1929 BP: 605 A; 385 C; 460 G; 479 T; 0 other:

Alignment Scores:

Pred. No.:	Length:	Score:
1.34e-50	1929	681.50
Percent Similarity:	Matches:	160
36.77%	Conservative:	72
Best Local Similarity:	Mismatches:	206
21.41%	Indels:	194
DB:	Gaps:	9

US-09-805-337A-2 (1-569) x AAS06029 (1-1929)

7 ValIleLeuIleSerTyrValSerThrValGlyGlyGluGlyThrLeuCysAspPhePro 26

63 ATTCTTTATTAATCTCTCAGAGAACTGTATGCAAGAGAAACAGTGTGATTTTCT 122

27 LysIleHisHisGly-----PheLeuTyrAsp---GluLysAspTyrAsnPro 41

123 ACCGTGAAATGAGAGAGATGCGCCATATTATATACGTTTAAAGCTTTACTCCCG 182

42 PheSerGluValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheValSer 61

183 ATGAGC-----GTAGACAAAAACATATCATCTCTGCTGCTGCTGCTATGCAACC 233

62 ProSerLysSerPheThrArgIleThrCysThrGluGluGlyThrSerProThrPro 81

234 GAAGTGGGAGAGCAAGAGAGCAATTCAGTGCACAGCAAGAGCTGTGTCCAAACCCA 293



XX Claim 1: SEQ ID NO 3840: 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumor in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 1315 BP; 454 A; 228 C; 238 G; 395 T; 0 other;

Alignment Scores:

pred. No.: 8.7e-51 Length: 1315  
Score: 681.00 Matches: 163  
Percent Similarity: 36.70% Conservative: 48  
Best Local Similarity: 28.35% Mismatches: 114  
Query Match: 21.39% Indels: 250  
DB: 24 Gaps: 11

US-09-805-337a-2 (1-569) x ABN97342 (1-1315)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTyrPheValSerThrValGlyIle---Glu 19  
DB 86 ATCTTTTACTATACATGATCGATTCGACCTTGCGTTCCCTGCTGAATGACACAGAA 145  
QY 20 GlyThrLeuCysAspPheProLysIleHisHisGlyPheLeuTyrAspGluGluAspTyr 39  
DB 146 GTGAAACCTGTGATTTTCCAGAAATTCACATGACGAGCTATATATAGAGTTGGCT 205  
QY 40 AsnProPheSerIleValProThrGlyGluValPheTyrTyrSerGlyTyrAsnPro 59  
DB 206 AGACTTACTCTTCACAGACGTCGACCAATCTTATCTTACGTGATCAAAATTTT 265  
QY 60 ValSerProSerLysSerPheTyrThrArgIleThrCysThrGluGluGlyTyrSerPro 79  
DB 266 GTGACGCTTCACAGAGTACGTGGATTCACATTCACGACACACAGATGGGTGGCA 325  
QY 80 ThrProLysCysLeuArgMetCysSerPhePro-----PheValLysAsnGlyHisSer 97  
DB 326 ACAGTCCCATGGCTTCAGAACATGCTCAAAATCAGATATAGAAATTCGAATTCAT 385  
QY 98 GluSerSerCysLeuIleHisLeuGluGlyAspThrValGluIleLeuLysAsnThrGly 117  
DB 386 TCTGAATCTTCTCTATTTATATTTTAATAAGAAATACAAATATTAATTAACACAGA 445  
QY 118 TyrSerLeuGln-----AsnAsnGluLysAsnIleSerCysValGluArgGlyTyrSer 135  
DB 446 TATGCAACACAGACAGATGGAATCTTCAGGTTCAATACATGTTGCAAAATGATGATGCA 505  
QY 136 ThrProProLysCys---SerPheThrLysGlyGluCysHisValProIleLeuGluAla 154  
DB 506 GCACAACCATTTGCGATTAATTT-----TGCGATATGCTGCTGTTTGGAG-- 550  
QY 155 AsnValAspAlaGlnProLysLysGluSerTyrLysValGlyAspValLeuLysPheSer 174  
DB 551 AATTCACAGACCCAGAGTAATGCGATGCGGTTTAAGCTCATGACACACATTGAGACTACGAA 610  
QY 175 CysArgLysAsnLeuIleArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyr 194  
DB 611 TGC----- 613  
QY 195 SerProAsnPheProThrCysLysGlyGlnValArgSerCysGlyProProProGlnLeu 214

DB 613 ----- 613  
QY 215 SerAsnGlyGluValLysGluIleArgLysGluGluTyrGlyHisAsnGluValGlu 234  
DB 613 ----- 613  
QY 235 TyrAspCysAsnProAsnPheIleIleAsnGlyProLysLysIleGlnCysValAspGly 254  
DB 614 TACGAT----- 619  
QY 255 GluTyrThrThrLeuProThrCysValGluGlnValLysThrCysGlyTyrIleProGlu 274  
DB 620 -----GGATAT-----GAA 628  
QY 275 LeuGluTyrGlyTyrValGlnProSerValProProTyrGlnHisGlyValSerValGlu 294  
DB 629 ATCAGTTATGCA----- 640  
QY 295 ValAsnCysArgAsnGluTyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGly 314  
DB 640 ----- 640  
QY 315 IleTyrThrGluLeuProMetCysValAlaThrHisGlnLeuLysArgCysLysIleAla 334  
DB 640 ----- 640  
QY 335 GlyValAsnIleLysThrLeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArg 354  
DB 640 ----- 640  
QY 355 IleArgTyrArgCysSerAspIlePheArgTyrArgHisSerValCysIleAsnGlyLys 374  
DB 640 ----- 640  
QY 375 TyrAsnProGluValAspCysThrGluLysArgGluGlnPheCysProProProGln 394  
DB 640 ----- 640  
QY 395 IleProAsnAlaGlnAsnMetThrThrThrValAsnTyrGlnAspGlyGluLysValAla 414  
DB 641 -----AACACACACAGCTCC----- 655  
QY 415 ValLeuCysLysGluAsnTyrLeuLeuProGluAlaLysGluIleValCysLysAspGly 434  
DB 656 -----ATAGGTGTGGTGAAGAT 673  
QY 435 ArgTyrGlnSerLeuProArgCysValGluSerThrAlaTyrCysGlyProProProSer 454  
DB 674 GGGTGGTCCCATTTTCCAAACATGTTATATTCTTCAGAAAGGTGGGCTCCCTCCACCT 733  
QY 455 IleAsnAsnGlyAspThrThrSerPheProLeuSerValTyrProProGlySerThrVal 474  
DB 734 ATTAGCAATGGATATACACCTCTTCTACTAAGATGTATGTGCGACAGTCAGAGATC 793  
QY 475 ThrTATGCGSelinSerPheTyrLysLeuGlnGlySerValThrValThrCysArgAsn 494  
DB 794 GAGTACCAATGACAGCTCTCTATGAACTTCACAGGTTCTATATATGTAACATGTAGTAAT 853  
QY 495 LysGlnTyrSerGluProProArgCysLysLeuAspProCysValLysGluGluAsnMet 514  
DB 854 GGAGAGTGTGGGAGACCCACAGATGCATGCATTCATTAATTAATTAATCTGAAGAAACATG 913  
QY 515 AsnLysAsnAsnIleGlnLeuLysTyrPArgAsnAspCylLysLeuTyrAlaLysThrGly 534  
DB 914 AATTAATAATATACATACGTTTAAAGGAAAGATGACATTAATAATATATGCAAAACAGGG 973  
QY 535 AspAlaValGluPheGlnCysLysPheProHisLysAlaMetIleSerSerProProPhe 554  
DB 974 GATACCACTGAATTTATGTATAATTGGGATTAATGGAATACGAAATACACAGTTATCATTT 1033  
QY 555 ArgAlaIleCysGlnGluGlyLysPheGluTyrProIleCysGlu 569  
DB 1034 CAAGCAGTGTGTAGGAGGACATAGTGGATTAACCCAGATCGCA 1078

```

RESULT 15
AAV02792
ID AAV02792 standard: DNA: 767 BP.
XX
XX AAV02792:
AC
XX
XX 27-APR-1998 (first entry)
XX
XX Human partial complement factor H cDNA fragment 2.
DE
XX
XX Complement factor H: tumour associated antigen; renal cancer;
KM urogenital cancer; medicament; modulator; ss.
XX
XX Homo sapiens.
OS
XX
XX MO9738136-A1.
PN
XX
XX 16-OCT-1997.
PD
XX
XX 09-APR-1997: 97MO-US05710.
PF
XX
XX 06-MAR-1997: 97US-0812481.
PR
XX 09-APR-1996: 96US-0015083.
PR 09-APR-1996: 96US-0630048.
PR 06-MAR-1997: 97US-0038614.
XX
XX (BARD-) BARD DIAGNOSTIC SCI INC.
PA
XX
XX Enfield DL, Hass GM, Kinders RJ;
PI
XX
XX WPI: 1997-512742/47.
DR
XX
XX P-PSDB: AAW39156.
XX
XX
XX Treating or screening for cancer, e.g. renal or urogenital cancer -
PT py modulating or detecting tumour associated human complement factor
XX H related antigen, or nucleic acid encoding it
XX
XX Example 6B: Fig 7A: 104pp: English.
XX
XX This partial cDNA sequence encodes a region of the human
XX tumour-associated complement factor H (CFH). This sequence is used
XX in the identification of DNA encoding complement factor H related
XX proteins and antigens from clone PBBS5FH2910 (see AAV02793-V02795). The
XX detection of this protein and a CFH antigen can be used in screening or
XX for the treatment of renal or urogenital cancer, e.g. bladder, cervical
XX or prostate cancer. Agents that may modulate this antigen could be used
XX in the manufacture of a medicament for the treatment of a tumour cell.
XX
XX
XX Sequence 767 BP; 257 A; 140 C; 179 G; 191 T; 0 other:
SO
Alignment Scores:
Pred. No.: 2,25e-36 Length: 767
Score: 515.00 Matches: 105
Percent Similarity: 47.47% Conservative: 45
Best Local Similarity: 33.23% Mismatches: 90
Query Match: 16.18% Indels: 76
DB: 18 Caps: 6
US-09-805-337A-2 (1-569) x AAV02792 (1-767)
OY 225 GIUGIUTYGYGLYHISASNGLUVALYALGIUTYRASPCYSASNPPOASNPHELLEIASN 244
DB 25 GACACTTACAGTATGAGAGAGAGAGTTCGTAACAATCTTTTGAAGCTTTTGAATTCAT 84
OY 245 GIYPTOLYSLYSILLEGINCYSVALISPGIYGIUTPTTHLTHLEUPROTHRCYSVALGIU 264
DB 85 GGGCCCTGCATTTGCAAAATGCTTAGAGAGAAAATGCTCACCTCCATCATGATTAATA 144
OY 265 GINVALYSLTHCYSGLYTYRIILEPROGLIULENGIUTYGYLYTYRVALGINSERVAL 284
DB 145 -----ACGATTTGCTCTAGTTACCTAGCTTGA-----AATGCCATA 183

```

```

OY 285 PRO-----PROTYGLINHSIGLYVALSERVALGLUVALASNCYSARG 298
DB 184 CCCATGGGAGAGAGAGAGATGATGTATAGCGCGGAGACAGTCACTTACACTTGTGCA 243
OY 299 ASNGIUTYRALMETLEGLYASNASMETLETHRCYSILIASNLYILETRPHRCIU 318
DB 244 ACATATTACAAAATGATGAGCCAGTAAATGTAACATGATTAATAGCAGATGACAGCA 303
OY 319 LEUPROMETCYSVALALETHRHSIGINLEULYSARGCYSYLSILEAGLYVALISNILE 338
DB 304 AGGCCACATGCG----- 315
OY 339 LYSTHLEULEULYSLEUSERGILYSGIUPHEASNHISASNSERARGILLEARGTYRARG 358
DB 315 ----- 315
OY 359 CYSSEASPILLEPHEARGTYRARGHISSEVALCYLSILEASNGILYSTTRPASNPPOGIU 378
DB 316 -----AGACACCTCTCTGTGTGAAT----- 336
OY 379 VALASPCYSTHRIULYSARGLUGLINPHECYSPPROPTROPROGLINLEPROASNALA 398
DB 337 -----CCGCCACAGTACAAATGCT 357
OY 399 GINASMETHTHTHTHYVALASN---TYRGINASPGIYGLIULYSVALALAVALLEUCYS 417
DB 358 TATATATAGTGTGACAGACAGATGATTAATCATCTGTGTGAGACAGATGATCATATCT 417
OY 418 LYSGLIASNTYRLEULEUPROGLIUALYSGIULILEVALCYSLYASASQIYARGTRPGIN 437
DB 418 AGGAGCCCTTATGAATGTTTGGGATGCAAGATGATGTGTTTAAATGGAAGACTGGACG 477
OY 438 SERLEUPROARGCYSVALGILUSERTHRALATYRCYSGIYPROPTROPROSERILLEASNASN 457
DB 478 GAACCACTCATGCAAAATGATTTCTAGAGAAATGCGGCCCTCCACACTTATGACAAT 537
OY 458 GLYASPTHTHTSERPHEPROLEUSERVALTYRPROPTROGLYSETHTHYVALTHTRYARG 477
DB 538 GGGGACATTACTTCCCTCCGTTGCAGTATATGCTCCAGCTTCAAGTTGACAGTCAAA 597
OY 478 CYSGLINSEPHETRYLISLEUGINGLYSERVALTHRYVALTHRCYSARGASNLGSHTRP 497
DB 598 TCCCAAGACTTATATCAACTTGAAGGTAAACAAGCAATATACATGTAGAAATGACAAATCG 657
OY 498 SERGIUPROPTROARGCYSLEUASPPROCYVALVALSERGLUGIULASNMETASNLYSASN 517
DB 658 TCAGAACCAACCAAAATGCTTACATCCGCTGTAATATATCCCGGAATATATGCAAAATTAT 717
OY 518 ASNILEGLINLEULYSTPRARGASNASPGIYLSLEUTYRALALYSTHR 533
DB 718 AACATAGCATTTAAGGTGGACAGCCCAAGAGAGAGCTTTATTCGAGACA 765

```

Search completed: July 23, 2003, 09:03:04  
Job time : 368 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: July 23, 2003, 03:08:43 ; Search time 2198 Seconds  
(without alignments)  
4192.553 Million cell updates/sec

Title: US-09-805-337A-2  
Perfect score: 3183  
Sequence: 1 MLTFSVILISWSTVGEG.....SSPFRALCOCKFEYFICE 569

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n model -DEV-rlh  
-O/cgn2.1/uspro.spool/us09805337/runat.17072003.135605.18928/app\_query.fasta.1.711  
-DB-EST -QPMF-fastap -SUFFIX-1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -DIST=45  
-DOCALLIGN-200 -THR\_SCORE-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTPMF-pio -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER-US09805337 @cgn2.1.1906 @cgnat.17072003.135605.18928 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLITLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -FGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estinu:\*  
4: em\_estlov:\*  
5: em\_estpl:\*  
6: em\_estro:\*  
7: em\_etc:\*  
8: em\_etc1:\*  
9: gdp\_est1:\*  
10: gdp\_est2:\*  
11: gdp\_etc1:\*  
12: gdp\_est3:\*  
13: gdp\_est4:\*  
14: gdp\_est5:\*  
15: em\_estfun:\*  
16: em\_estlom:\*  
17: gdp\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_liv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2384	74.9	1411	11	BC026282
2	1037	32.6	683	12	BG616762
3	1016	31.9	646	10	AV657525
4	1014	31.9	633	10	AV657514
5	828.5	26.0	795	10	AV683459
6	801	25.2	832	12	BG562869
7	797.5	25.1	858	12	BG569423
8	792.5	24.9	853	9	AU138301
9	788.5	24.8	912	12	BG564962
10	785.5	24.7	786	12	BG533258
11	780	24.5	839	12	BG533248
12	777	24.4	790	12	BG286384
13	777	24.4	825	10	AV682232
14	764	24.0	761	10	AV688058
15	760.5	23.9	735	12	BG616866
16	759.5	23.9	787	12	BG563052
17	759	23.8	633	10	AV649114
18	758	23.8	723	10	AV682522
19	757.5	23.8	846	12	BG616668
20	757	23.8	739	12	BG563738
21	753	23.7	801	12	BG563289
22	749	23.5	675	13	B1254780
23	745.5	23.4	823	12	BG563035
24	742.5	23.3	777	12	BG567040
25	739.5	23.2	838	12	BG617334
26	735	23.1	721	12	BG568304
27	727.5	22.9	650	12	BG565462
28	726	22.8	744	12	BG563564
29	724.5	22.8	739	9	AU139391
30	721.5	22.7	681	12	BG568358
31	721.5	22.7	780	12	BG565469
32	720	22.6	1113	12	BG567988
33	719	22.6	668	10	AV650989
34	718.5	22.6	739	12	BG617078
35	716.5	22.5	886	12	BG568871
36	715.5	22.5	771	10	AV682102
37	715.5	22.5	777	12	BG618219
38	714	22.4	785	12	BG569316
39	714	22.4	800	12	BG568888
40	713.5	22.4	836	12	BG564780
41	708.5	22.3	724	12	BG617992
42	708.5	22.3	880	12	BG569198
43	705	22.1	643	10	AV681962
44	703.5	22.1	802	12	BG617178
45	701.5	22.0	617	12	BG563181

## ALIGNMENTS

RESULT 1  
BC026282  
LOCUS  
DEFINITION  
Homo sapiens, Similar to factor H-related protein 5, clone IMAGE:4733780, mRNA.  
ACCESSION  
BC026282  
VERSION  
BC026282.1 GI:20070808  
KEYWORDS  
HTC.  
SOURCE  
Homo sapiens.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1411)  
AUTHORS  
Strausberg, R.  
TITLE  
Direct Submission

## JOURNAL

Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabps-r@mail.nih.gov](mailto:cgabps-r@mail.nih.gov)  
Tissue Procurement: CLONTECH  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mdedpaxi@stanford.edu](mailto:mdedpaxi@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRL Plate: 37 Row: 1 Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13540562  
This clone has the following problem: frame shifted.

## FEATURES

## source

1..1411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4733780"  
/tissue\_type="Liver"  
/clone\_lib="NHL MGC\_76"  
/lab\_host="DH10B"  
/note="Vector: pDNR-LIB"

BASE COUNT 496 a 261 c 275 g 379 t  
ORIGIN

## Alignment Scores:

Pred. No.: 6 45e-247 Length: 1411  
Score: 2384.00 Matches: 427  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.77% Mismatches: 0  
Query Match: 74.90% Indels: 0  
DB: 11 Gaps: 0

US-09-805-337a-2 (1-569) x BC026282 (1-1411)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
DB 110 ATGTTGCTCTTATTCAGTGAATCCTAATCTCATGGTATCCAGTGTGGGAGAGAGA 169  
QY 21 ThrLeuGlyAspPheProIlysiIleHisGlyPheLeuTrpAspGluLysPtyrAsn 40  
DB 170 ACACATTGGATTTCACAAAATACACACATGATTTCTGATGATGAGAAAGATTATAAC 229  
QY 41 ProPheSerGlnValProThrGlyGluValAlaPheTrpTrpSerCysGluTrpAsnPheVal 60  
DB 230 CCTTTTCCCAAGTCTCTACAGGGAGATTTCTATTACTCCGTGAATAATATTTTCGG 289  
QY 61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr 80  
DB 290 TCTCCTCAAAATCCTTTTGACCTCGCATACATGACAGAAAGAGATGTCACCAACA 349  
QY 81 ProLysGlyLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer 100  
DB 350 CCGAAGTGTCTCAGATGTCTTCTTCTTTGTCAAAAAAGCATTCATCTGAATCTTCA 409  
QY 101 GlyLeuIleHisLeuGluLysAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu 120  
DB 410 GGACTAATACATCTGGAGGCTATCTGACAAATATTATTGACACACAGATACAGCTT 469  
QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProIleCys 140  
DB 470 CAAACAAATGACAAAAACATTTCTGTGTGTAGAACGGGCTGTGCTCCACTCTCCCATATTC 529

QY 141 SerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160  
DB 530 AGCTTCATTAAGGACAAATGTATGTCCAAATTTTGAAGCCCAATGTATCATGCTACGCCA 589  
QY 161 LysLysGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180  
DB 590 AAAAAAGAAAGCTACAAAGTTGGACGCTGTGAATTCCTCTCGCAAAAAATCTTATA 649  
QY 181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr 200  
DB 650 AGAGTTGGATCAGACACCTCAATTCATATTTACCAATTTGGTGTGCTACCTTATTCACACA 709  
QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyLysValLys 220  
DB 710 TCGAAAGACAGATGATCATGTGTGTGCTCCACTCTCCCACTCTCCCAATGTGAAGTTAG 769  
QY 221 GluIleArgLysGluLysTrpGlyHisAsnGluValGluTrpAspCysAsnProAsn 240  
DB 770 GAGATAGAAAGAGAGAAATATGACACAAATGATGATGATGATGATGATGATGATGAT 829  
QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyLysTrpThrLeuPro 260  
DB 830 TTTATATATAACGGGCTTAAAGAAATACATGTGTGATGACGAATGACAACTTACCC 889  
QY 261 ThrCysValGluGluValLysThrCysGlyTrpIleProGluLeuGluTrpGlyTrpVal 280  
DB 890 ACTTGTGTGACAAACATGAAACATGTGATACATCTGACATCTGACATCTGATCTTATGTT 949  
QY 281 GlnProSerValProProTrpGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300  
DB 950 CAGCCCTGTGCTCCCTCATCAACATGAGATTTGACGTGACGTGATTCAGAAATGAA 1009  
QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro 320  
DB 1010 TATGCAATGATTTGAAATATACATGATGATGATGATGATGATGATGATGATGATGAT 1069  
QY 321 MetCysValAlaThrHisGlnLeuLysAspGlyLysIleAlaGlyLysAsnIleLysThr 340  
DB 1070 ATGTGTGTGACAAACACCACTTAAAGAGCTCAAAATAGCAGAGTTTAAATTAAGAAACA 1129  
QY 341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer 360  
DB 1130 TTACTCAACCTATCTGGGAAAGAAATTTAATCAATTTCTAGAAATAGTTACAGATGTCA 1189  
QY 361 AspIlePheArgTrpArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp 380  
DB 1190 GACATCTTCAGATACAGCAGCTCAGTGTATTAACGGGAATGGAATCTGACATGAC 1249  
QY 381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400  
DB 1250 TGCACAGAAAAAGGAAACAAATTCGCCACGCCACCTCAGATACCTAATCTCAGAAAT 1309  
QY 401 MetThrTrpThrValAsnTrpGlnAspGlyGluLysValAlaValLeuLysGluLys 420  
DB 1310 ATGACACACAGTGAATATATGAGATGAGAGAAAAAGTAGCTGTCTCTGTAAACAAAC 1369  
QY 421 TyrLeuLeuProGluAlaLysGlu 428  
DB 1370 TATCTACTTCCAGAAAGCAAAAAA 1393

RESULT 2  
Bg616762 683 bp mRNA linear EST 18-APR-2001  
LOCUS 602614987F1 NTH\_MGC\_76 Homo sapiens CDNA clone IMAGE:4733780 5',  
DEFINITION  
mRNA sequence.  
ACCESSION Bg616762  
VERSION Bg616762.1 GI:13668133  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 683)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@email.nih.gov](mailto:cgabbs@email.nih.gov)  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LCM1595 row: n column: 21  
 High quality sequence stop: 683.

## FEATURES

source  
 Location/Qualifiers  
 1.683  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NIH-MGC-76"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: liver; Vector: pDNR-LIB (Clontech). Site\_1: SfiI (ggccgctcgcc); Site\_2: SfiI (ggccatagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATATAGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 220 a 132 c 130 g 201 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1,35e-101 Length: 683  
 Score: 1037.00 Matches: 190  
 Percent Similarity: 99.48% Conservative: 0  
 Best Local Similarity: 99.48% Mismatches: 0  
 Query Match: 32.58% Indels: 1  
 DB: 12 Gaps: 0

US-09-805-337A-2 (1-569) x BG616762 (1-683)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
 Db 111 ATGTTGCTCTATTCAGTGAATCCTAATCTCATGCGTATCCACTGTGGGAGAGGA 170  
 QY 21 ThrLeuCysAspPheProIysIleHisIsglyPheLeuTyraSpGluAspTyra 40  
 Db 171 ACACCTTGATGATTTCCAAAATAACACCATGATTTCTGTATGATGAGAAATATTAAC 230  
 QY 41 ProPheSerGlnValProThrGlyGluValPheTyrtysSerGlyValAsnPheVal 60  
 Db 231 CCGTTTCCCAAGTCTCTACAGGGAAGTTTCTATTAACCCGTGAATAATTAATTCGTG 290  
 QY 61 SerProSerIysSerPheTrpTrpArgIleThrCysThrGluGluGlyTrpSerProthr 80  
 Db 291 TGTCCCTCAAAATCCCTTGGACTCCCATACATGCACAGAAAGAGATGTCACCAACA 350  
 QY 81 ProIysCysLeuArgMetCysSerPheProPheValIysAsnGlyHisSerGlnSer 100  
 Db 351 CCGAAGTGTCTCAGATGTGTTCCCTTGTGAAAAATGTCATTCGTAATCTTCA 410  
 QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleLeuGlyAsnThrGlyTrpSerLeu 120  
 Db 411 GGCATTATTCATCTGAGAGGTGATCTGACAAATATTTGACACAGCATACACCCCTT 470  
 QY 121 GlnAsnAsnGluIysAsnIleSerCysValGluArgGlyTrpSerThrProPheIleCys 140  
 Db 471 CAAACAATAGAGAAAAATTTGCTGTGTGAACGGGGCTGCTCCACTCTCCCATATGC 530  
 QY 141 SerPheThrIysGlyIleCysHisValProIleLeuGluValAsnValAspAlaGlnPro 160

Db 531 ACCTGACTAAAGAGAGATGTATGTTCCAAATTTTGAAGCCATGAGATGCTACGCCA 590  
 QY 161 LysLysGlnSerTrpIlyValGlyAspValLeuIlyPheSerCysArgIysAsnLeuIle 180  
 Db 591 AAAAAAGAAAGCTACCAAGTGGAGACGTGTGAATTTCTCTGCAGAAAAAATCTTATA 650  
 QY 181 ArgValGlySerAspSerValGlnCysTy 190  
 Db 651 CAGAGTTGATTCAGACTCACTTCAATGTTAC 681

RESULT 3  
 AV657525 646 bp mRNA linear EST 16-JAN-2002  
 LOCUS AV657525  
 DEFINITION AV657525 GLC Homo sapiens cDNA clone GLCFCG11 3', mRNA sequence.  
 ACCESSION AV657525  
 VERSION AV657525.1 GI:9878539  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 646)

## AUTHORS

Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
 Xiao,H., Lu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,  
 Shen,K., Ou,J., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,  
 Hu,G., Gu,J., Chen,Z. and Han,Z.  
 Insight into hepatocellular carcinogenesis at transcriptome level  
 by comparing gene expression profiles of hepatocellular carcinoma  
 with those of corresponding noncancerous liver  
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
 21625106

## JOURNAL

## MEDLINE

Contact: Zenguan Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801919(ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@chgc.sh.cn  
 This clone is available at CHGC in Shanghai.

## FEATURES

source Location/Qualifiers

1.646

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="GLCFCG11"

/clone\_lib="GLC"

/tissue\_type="corresponding non cancerous liver tissue"

/dev\_stage="Adult"

/lab\_host="SOLAR"

/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 201 a 132 c 121 g 190 t 2 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.32e-99 Length: 646  
 Score: 1016.00 Matches: 188  
 Percent Similarity: 87.56% Conservative: 2  
 Best Local Similarity: 86.64% Mismatches: 8  
 Query Match: 31.92% Indels: 19  
 DB: 10 Gaps: 1

US-09-805-337A-2 (1-569) x AV657525 (1-646)

QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
 Db 46 ATGTTGCTCTATTCAGTGAATCCTAATCTCATGCGTATCCACTGTGGGAGAGGA 105  
 QY 21 ThrLeuCysAspPheProIysIleHisIsglyPheLeuTyraSpGluAspTyra 40  
 Db 106 ACACCTTGATGATTTCCAAAATAACACCATGATTTCTGTATGATGAGAAATATTAAC 165  
 QY 41 ProPheSerGlnValProThrGlyGluValPheTyrtysSerGlyValAsnPheVal 60

Db		166	CCTTTTCCCAAGTTCCTACAGGGGAATTTCATTAATCTCCTGTGATAAATTTCTGTG	225
Oy		61	SerProSerIysSerPheTrpThrArgIleThrCysThrGluGlyTyrSerP	80
Db		226	TCTCCTCCAATCCTTTTGACTCGCATACATGCACAGAAGATGCTCACAAACA	285
Oy		81	ProLysCysLeuAtgMetCysSerPheProPheValLysAsnGlyHisSerGluSer	100
Db		286	CCGAAGTCTCTCAGAAATGTCTTCTCTTCTTTGTGAAAATGTCATCTGMAATCTTCA	345
Oy		101	GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnTrpGlyTyrSerLeu	120
Db		346	GGACTAAATACATCTCGAAGTGAAGTGAACGTACAAATTAATTTGCCAACACAGATACAGCCTT	405
Oy		121	GlnAsnAsnGlnLysAsnIleSerCysValGlnArgGlyTyrSerThrProProIleCys	140
Db		406	CAAAACAAATGAGAAAAACATTTCGTGTGTATACACCGNCTGTGTCACCTCCCAATATGC	465
Oy		141	SerPheThrIysGlyGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro	160
Db		466	AGCTTCACTAAGAGAGATGTCATGTTCCATTCATTAGAAAGCAAATGTAGATGCTCAGCA	525
Oy		161	LysIysGluSerTyrIlysValGlyAspValLeuLysPheSerCysArg-LysAsnLeuIl	180
Db		526	AAAAAAGAAACCTACACAGTGTGGAGACGTGTGAATTCCTCTCGAGATCA-----	577
Oy		180	eArGVaIGlySeraspSerValGlnCysTyGIlnPheGlyTrpSerProAsnPheProTh	200
Db		578	-----CCTAACTTTTCAAC	591
Oy		200	rCyLySGlyGlnValArgSerCysGlyProProProGlnLeuSerAsn	216
Db		592	ATGCAMAGACAACTACGACACATGTGTGTCACCTTCCTCAATCTTCAAT	640
<hr/>				
LOCUS	AV657514	633 bp	mRNA	linear EST 16-JAN-2002
DEFINITION	AV657514	GLC Homo sapiens cDNA clone GLOCFC12 3', mRNA sequence.		
ACCESSION	AV657514			
VERSION	AV657514.1	GI:9878528		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	human.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
	1 (bases 1 to 633)			
	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,			
	Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,			
	Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,			
	Hu,G., Gu,J., Chen,Z. and Han,Z.			
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level			
	by comparing gene expression profiles of hepatocellular carcinoma			
	with those of corresponding noncancerous liver			
	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)			
JOURNAL MEDLINE COMMENT	Contact: Zeguang Han			
	Chinese National Human Genome Center at Shanghai			
	351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai			
	201203, P. R. China			
	Tel.: 86-21-50801919(ex. 45)			
	Fax: 86-21-50801922			
	Email: hanzg@chgc.sh.cn			
FEATURES	This clone is available at CHOC in Shanghai.			
SOURCE	Location/Qualifiers			
	1..633			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="GLCF12"			
	/clone_lib="GLC"			
	/tissue_type="corresponding non cancerous liver tissue"			
	/dev_stage="Adult"			
	/lab_host="SOLR"			

BASE COUNT	193 a	132 c	123 g	185 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3.7e-99		Length:	633
Score:	1014.00		Matches:	186
Percent Similarity:	87.85%		Conservative:	2
Best Local Similarity:	86.92%		Mismatches:	8
Query Match:	51.86%		Indels:	18
DB:	10		Gaps:	1
US-09-805-337A-2 (1-569) x AV657514 (1-633)				
QY	1 MetleuleuLeuPheSerValIleIleuLeuIleSerThrPValSerThrValIGlyGlyGly	20		
DB	46 ARGTGGCTCTTAATCAGTATACGTATACCTTAATCAGTATACGATGGATACGACGTTGGGGACAACAGA	105		
QY	21 ThrIeuCysAspPheProIysIleHisIsglyPheLeuIYrAspGluGluAspTryAsn	40		
DB	106 ACACCTTGTCGATTTCCCAAAAATACACCATGGATTTCTGTATMGATGAGAGAAATATATTAAC	165		
QY	41 ProPheSerGluValProThrGlyGluValPheYrYrYrSerCysGluTyAsnPheVal	60		
DB	166 CCTTTTTCACAAAGTCTTCACAGGGAGAAATTTCTATATCTCTCGATGAATATATTTTGTG	225		
QY	61 SerProSerLysSerPheTrpThrArgIleThrCysThrGluGluGlyTyrPserProThr	80		
DB	226 TCTCTCTCAAAATCTCTTTTGGACATCGCATTAACATGCACAGAAAGATGGTCACACACA	285		
QY	81 ProIYcysLysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSerSer	100		
DB	286 CCGAAGTGTCTCAGATATGTCTTCTCTTCTTTTGAAAATGTGCATTCGAATCTTCA	345		
QY	101 GlyIleuIleHisIleGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu	120		
DB	346 GGACTAATACATCTCGAAGGTGATACGTACAAATATATTTCCACACACGATACAGCCTT	405		
QY	121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTyrPserThrProProIleCys	140		
DB	406 CAAAACATGAGAAAAACATTCGCTGTACACAGCGGCTGTCACATCCCTCCATATGC	465		
QY	141 SerPheTrpLysGlyGluCysHisValProIleIleuGlnIleAsnValAspAlaGlnPro	160		
DB	466 AGCGTTACTTAAGGGAATGTCATGTTCCAAATTTTACAGCCAAATGATGATGCTCAGACA	525		
QY	161 LysLysGluSerTrpLysValGlyAspValIleuLysPheSerCysArgLysAsnIleuIle	180		
DB	526 AAACAAGAGACTACACAGATGCAGAGACGTTGCAATCTCTCTGACAG	573		
QY	181 ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTyrPserProAsnPheProThr	200		
DB	574 -----ACACCTTAATCTTTTCACAA	591		
QY	201 CysLysGlyGlnValArgSerCysGlyProProGlnIleu	214		
DB	592 TGCACAGCACAAGTACGATCATGTGTGGCCACCCTCTTCATTT	633		
RESULT 5				
LOCUS	AV683459	795 bp	mRNA	linear EST 16-JAN-2002
DEFINITION	AV683459 GKC Homo sapiens cdna clone GKDDP09 5', mRNA sequence.			
ACCESSION	AV683459			
VERSION	AV683459.1	GI:10285322		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,			





RESULT 6					
LOCUS	BC562869				
DEFINITION	BC562869	832 bp	mRNA	linear	EST 10-APR-2001
	602581712r1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4709426 5',				
ACCESSION	BC562869				
VERSION	BC562869.1	GI:13570521			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 832)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D.				

FEATURES	Location/Qualifiers
SOURCE	1..832

BASE COUNT	275 a	171 c	154 g	232
ORIGIN				

Alignment Scores:	
Pred. No.:	6,31e-76
Score:	801.00
Percent Similarity:	36.83%
Best Local Similarity:	32.97%
Query Match:	25.16%
DB:	12
	Gaps: 3
	Mismatches: 39
	Conservative: 21
	Matches: 179
	Indels: 304
	Length: 832

US-09-805-337A-2 (1-569) x BG562869 (1-832)

[illegible]

QY	8	ProLysGlnLeuArgMetCysSerPheProPheValLysAsnGlnYHISerGluSer	100
Db	323	CCAAAGGCTCGACCTGCTGTTCTTCTCTTCTTGGAAGGATGTCGAAATCTCA	362
QY	101	GlyLeuLeuHisLeuGlnGluLysAspThrValGlnIleIleCysAsnThrGlyTrpSerLeu	120
Db	383	GGACAAACACATCTGGAGGTGATCTGTACCAAAATTATTGCAACACAGCATACAGACTT	442
QY	121	GlnAsnAsnGlnLysAsnIleSerCysValGluArgGlyTrpSerThrProProIleCys	140
Db	443	CAAAACATATAGAAACAACATTTTCATGTGTAGAAACGGGCTGTGTACACCTCCCAAAATGC	502
QY	141	SerPheThrLysGlnGluCysHisValProIleLeuGlnAlaAsnValAspAlaGlnPro	160
Db	502	-----	502
QY	161	LysLysGlnSerThrLysValGlnLysPheLeuLysPheSerCysArgLysAsnLeuIle	180
Db	502	-----	502
QY	181	ArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProThr	200
Db	502	-----	502
QY	201	CysLysGlnGlnValArgSerCysGlyProProProGlnLeuSerAsnGlnGluValLys	220
Db	502	-----	502
QY	221	GluIleArgGlySerGluGluTrpGlnHisAsnGluValIleGluTrpAspCysAsnProAsn	240
Db	502	-----	502
QY	241	PheIleLeuAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro	260
Db	502	-----	502
QY	261	ThrCysValGlnGlnValLysThrCysGlyTyrlleProGluLeuGluTrpGlyTyrlVal	280
Db	502	-----	502
QY	281	GlnProSerValProProTrpGlnHisGlyValSerValGluValAsnCysArgAsnGlu	300
Db	502	-----	502
QY	301	TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPro	320
Db	502	-----	502
QY	321	MetCysValAlaThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr	340
Db	502	-----	502
QY	341	LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTrpArgCysSer	360
Db	502	-----	502
QY	361	AspIlePheArgTrpArgHisSerValCysIleAsnGlyLysTrpAsnProGluValAsp	380
Db	502	-----	502
QY	381	CysThrGlnLysArgGlnGlnPheCysProProProProGlnIleProAsnAlaGlnAsn	400
Db	502	-----	502
QY	401	MetThrThrThrValAsnTrpGlnAspGlyGlnLysValAlaIleValLeuCysLysGluAsn	420
Db	502	-----	502
QY	421	TyrLeuLeuProGlnAlaLysGluIleValCysLysAspGlyArgTrpGlnSerLeuPro	440
Db	502	-----	502

QY	441	ArgcysValIuSerThrAlaTyrCysGlyProProserIleAsnAsnGlyAspThr	460
Db	503	AGGTCACCTATTTCGTGCAGAAAAATGTGGGCCCTCCACTATTGCACATGGAGACATT	562
QY	461	ThrSerPheProLeuSerValTyrProProGlySerThrValThyTyrArgCysGlnSer	480
Db	563	ACCTTCATCCCTGGTTGTCAGTATATGCTCCAGAGTTTCATCAGTTGATGCATCCATGCCAGAAC	622
QY	481	PhetyrIyrIeuGlnGlySerValThrValThrcysArgAsnIysGlnTrpSerGluPro	500
Db	623	TTGTATCAACTGTGAGGGTAACATCAATAATACATGTAAGAAACGGACAAATGGTCAGAAACCA	682
QY	501	ProArgCysLeuAspProCysValValSerGluIuAsnMet-AsnIysAsn---AsnI	519
Db	683	CCAAATGCTTAGATTCATGATGTATATTCACCAAGAAATTCATGCAACATTTATACAT	742
QY	519	egInLeuIys---TrrArgAsnAspGlyIysLeuTyrAla-LysThrGlyAspAlaValG	538
Db	743	AAATTTCAAGTGGGACAAAACAACAAAAGCTTTATTTCACGACGATGACATAGTTG	802
QY	538	IuPhe	539
Db	803	AATTC	807
RESULT 7	803	AATTC	807
LOCUS	BG569423	858 bp	linear
DEFINITION	602588766F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722761 5',		
ACCESSION	BG569423		
VERSION	BG569423.1	GI:13577076	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
LOCUS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
DEFINITION	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
ACCESSION	1 (bases 1 to 858)		
VERSION	NIH-MGC Htlcp://mgc.nci.nih.gov/.		
KEYWORDS	National Institutes of Health, Mammalian Gene Collection (MGC)		
SOURCE	Unpublished (1999)		
ORGANISM	Contact: Robert Strausberg, Ph.D.		
DEFINITION	Email: cga@brrfemail.nih.gov		
ACCESSION	Tissue Procurement: CLONTECH Laboratories, Inc.		
VERSION	cDNA Library Preparation: CLONTECH Laboratories, Inc.		
KEYWORDS	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
SOURCE	DNA sequencing by: Incyte Genomics, Inc.		
ORGANISM	Clone distribution: MGC clone distribution information can be		
DEFINITION	found through the I.M.A.G.E. Consortium/LNL at:		
ACCESSION	http://image.llnl.gov		
VERSION	Plate: LICM1581 row: c column: 18		
KEYWORDS	High quality sequence stop: 758.		
SOURCE	Location/Qualifiers		
FEATURES	1. 858		
FEATURES	/organism="Homo sapiens"		
FEATURES	/db_xref="taxon:9606"		
FEATURES	/clone="IMAGE:4722761"		
FEATURES	/clone_lib="NIH_MGC_76"		
FEATURES	/lab_host="DH10B (TI phage-resistant)"		
FEATURES	/note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site_1:		
FEATURES	SfiI (ggccgcctcgcc); Site_2: SfiI (ggccatgtgccc); 5' and		
FEATURES	3' adaptors were used in cloning as follows: 5' adaptor		
FEATURES	sequence: 5'-CACGGCCATATGGCC-3' and 3' adaptor		
FEATURES	sequence: 5'-ATTCTAGAGCCGACGGCCGACGATG-dT(30)BN-3' (where B = A,		
FEATURES	C, or G and N = A, C, G, or T). Average insert size 1.8		
FEATURES	kb (range 1.0-4.0 kb). 15/15 colonies contained inserts		
FEATURES	by PCR. This library was enriched for full-length clones		
FEATURES	and was constructed by Clontech Laboratories (Palo Alto,		
FEATURES	CA). Note: this is a NIH-MGC library."		
BASE COUNT	276 a	181 c	157 g
ORIGIN		244 t	
Alignment Scores:	1,58e-75	Length:	858
Prod. No.:			

Score:	797.50	Matches:	183
Percent Similarity:	36.44%	Conservative:	24
Best Local Similarity:	32.22%	Mismatches:	56
Query Match:	25.0%	Indels:	306
DB:	12	Gaps:	2
US-09-805-137A-2 (1-569) x BG564923 (1-858)			
QY	1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20		
DB	69 ATGTGGCTCTGCTGCTAGTAACTTAATCTACCGAATATCTGTGGGGGAGGCA 128		
QY	21 ThrLeuCysAspPheProLysIleHisHisGlyPheLeuTrpAspGluGluAspTrpAsn 40		
DB	129 ATGTCTGTGATTTTCCAAAATAAACCTGGAAATTCATATGATGAGAAATAATATTAAG 188		
QY	41 ProPheSerGlnValProThrGlyGluValPheTrpTrpSerCysGluTrpAsnPheVal 60		
DB	189 CCAATTTTCCAAATCTCTACAGGGGAAGTTTCTATTACTCTGTGAATATTAATTTGTG 248		
QY	61 SerProSerTrpSerPheTrpThrArgIleThrCysThrGluGluGlyTrpSerProThr 80		
DB	249 TGTCTCTCAAAATCCTTTTGGACGTGCATAACGTGCGAGAAAGGATGGTCAACACA 308		
QY	81 ProLysCysLeuArgMetCysSerPheProPheVal-LysAsnIleHisSerGluSer 100		
DB	309 CCAAGTGTCTGACTGTGTTCTTCTCTTTTGGAAACATGTGATTTCTGAACTTC 368		
QY	100 rGlyLeuIleHisLeuGluGlyAspThrValGlnIleIleLysAsnThrGlyTrpSerLe 120		
DB	369 AGGACAAACACATCTGGAAGGTGAATCTGTACAAATTTATTGGCAACACAGATTCAGACT 428		
QY	120 uGlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTrpSerThrProPheLeu 140		
DB	429 TCAAAACAATGAGAACACATTTCAATGATGTATGAAAGGGGCTGTCCATCTCCCAATG 488		
QY	140 sSerPheThrLysGlyGluCysHisValProIleLeuGluAlaAsnValAspAlaGlnPr 160		
DB	489 C----- 489		
QY	160 oLysGluGluSerTrpLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIl 180		
DB	489 ----- 489		
QY	180 eArgValGlySerAspSerValGlnCysTrpGlnPheGlyTrpSerProAsnPheProth 200		
DB	489 ----- 489		
QY	200 rGlyLysGlyGlnValArgSerCysGlyProProGluLeuSerAsnGlyGluValLy 220		
DB	489 ----- 489		
QY	220 sGluIleArgLysGluGluTrpGlyHisAsnGluValGluGlyTrpAspCysAsnProAs 240		
DB	489 ----- 489		
QY	240 nPheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrTrpLeuPr 260		
DB	489 ----- 489		
QY	260 oThrCysValGluGlnValLysThrCysGlyTrpIleProGluLeuGluTrpGlyTrpVa 280		
DB	489 ----- 489		
QY	300 uTrpAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuPr 320		
DB	489 ----- 489		
QY	320 oMetCysValAlaTrpHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysTh 340		

```

Db      489 .....
QY      340 rleuleuylsleuserglylsylglupheasnhsanserarglleargtyrarcysse 360
Db      489 .....
QY      360 rAspIlepheargtyrarghisservalcyslleasnilylstrpsnprogluvalas 380
Db      489 .....
QY      380 pCysThrGluLysArgGluInPheCysProProProGlnIleProAsnIacIlnas 400
Db      489 .....
QY      400 mMetThrThrThrValAsnTyrglnaspGlyGluLysValAlaValLeuCysLysGlnas 420
Db      489 .....
QY      420 nTyrlleuleuprogluvalalysgluilevalcyslysaspGlyArgTrpGlnSerleupr 440
Db      489 .....
QY      440 oArgCysValGluSerThrAlaTyrcysGlyProProProSerIleAsnAsnGlyAspTh 460
Db      490 -AGGTCCACTATTCTGCAGAAAAAATGTGGCCCTCCACCTCATTTGACACATGAGACAT 548
QY      460 rThrSerPheProLeuSerValTyrrProGlySerThrValThrTyraGcysGlnse 480
Db      549 TACTTCATTCCTGCTGTGTCTGATATAGCTCCAGCTTCATCAGTTGATGACAGTGCAGACA 608
QY      609 CTGTGTCAAACTTGAGGTGAGGTACCAATCAATCAATGAGCAACGCAATGCTGAGAAC 668
Db      500 rCPProArgCysLeuAspProCysValAlaSerGluGlnLysMetLysnLysAsnIleG 520
QY      669 CACCAAAATGCTTAGTTCATGTTATATCCCAAAATTTATGCAACATATACCTAA 728
Db      520 InLeuLysTrpArgAsnAspGlyLysLeuTyraLysThrGlyAspAlaValAlaGluPheG 540
QY      729 AATTACAGTGGACAAAGCAACAAAGCTTATCC-AAGACAGTGCATATTGACTTGG 787
Db      540 IncysLysPheProHisLysAlaMetIleSerSerProProPheargAlaIleCysGlnG 560
QY      788 TTCGTAACCTCCGAATCATATCCCAAAATCTCA-----TTCAATTGACGATGTTTGG 838
Db      560 lGcLysPheGluTyrrPro 566
QY      839 AATTGAAACTGTATTTCCCG 858

```

RESULT 8  
 AU138301 853 bp mRNA linear EST 02-AUG-2002  
 LOCUS AU138301  
 DEFINITION AU138301 PLACel Homo sapiens cDNA clone PLACE1008277 5', mRNA  
 sequence.  
 ACCESSION AU138301  
 VERSION AU138301.1 GI:10999822  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 853)  
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
 Isogai,T.  
 HRI human cDNA project  
 Unpublished (2000)  
 Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
 Research Institute; cDNA library construction; Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers

FEATURES  
 source  
 1. 853  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1008277"  
 /clone\_lib="PLACel"  
 /issue\_type="placenta"  
 /note="Vector: pME18FPL3"  
 BASE COUNT 297 a 152 c 185 g 214 t 5 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,45e-75 Length: 853  
 Score: 792.50 Matches: 143  
 Best Local Similarity: 72.848 Conservative: 34  
 Best Local Similarity: 58.858 Mismatches: 54  
 Query Match: 24.908 Indels: 12  
 DB: 9 Gaps: 3

US-09-805-337a-2 (1-569) x AU138301 (1-853)

```

QY      226 GUTYRGYlHnAsnGluValAlaGluTyraPcysAsnProAsnPhelIleasnGly 245
Db      2 GAATATGACACAGGAGTGTGATGATTTATTTGCAATTCAGATTCTTAATGAGAGGA 61
QY      246 ProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuProThCysValGluIn 265
Db      62 CCTAATAAATTCAGTGCTTGTATGAGAGAGTGCAGACACTTTCACAGTGTGTGCGAG 121
QY      266 ValLysThrCysGlyTyrrleProGluLeuGluTyrglyTyraGlnProserValPro 285
Db      122 GAGAGTACCTGCGAGATATACCTGCAACTTGAACATGGCTGGCCAGCTTTCCTCCCT 181
QY      286 ProTyrglnHisGluValSerValGluValAsnCysArgAsnGluTyraLametiLeGly 305
Db      182 CCTTATTAATGAGATTCAGTGCATTCGAATTCATGCTGAGAAATCAATTAATGATTTGGA 241
QY      306 AsnAsnMetIleThrCysIleAsnGlyIleTrpThrGluLeuProMetCysValAlaThr 325
Db      242 CACAGATCAATTACGATGATTCATGAGATGAGCAGCAACTTCCCAGTGTGCGCAATTA 301
QY      326 HisGlnLeuLysArgCysLys-----IleAlaGlyValAsnIleLysThr 340
Db      302 GATAAACCTTAAGAGAGTGCAGAAATCATCAAAATTTAATTACTTGGGGACAAATTTAAAA 358
QY      341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyraGcysSer 360
Db      359 -----AACAGAGGAATTCGATCTAATATCTAACATTAAGGACGATGATGACA 406
QY      361 AspIlepheargtyrarghisservalcyslleasnilylstrpsnprogluvalasp 380
Db      407 CGAAAAAGAGATGATGATACACACACAGCTGCGCAATTAATGCAAGATGGCATCCAGTGAAC 466
QY      381 CysThrGluLysArgGluInPheCysProProProGlnIleProAsnIacIlnas 400
Db      467 TGCTCAATGCGCAAAATCAATTAATGAGCCACCTCCACCTCAGATTTCCCAATTTCTCAAT 526
QY      401 MetThrThrThrValAsnTyrglnaspGlyGluLysValAlaValLeuCys-LysGlnas 420
Db      527 ATGACACACACACTGAATTAATGCGAGTGGAGAGAAAGATATCTTGTCCAAAGAAA 586
QY      420 nTyrlleuleuprogluvalalysgluilevalcyslysaspGlyArgTrpGlnSerleupr 440
Db      587 TTATCTAATTGCGAAGAGCAAGAAATTAATGCAAGATGAGNAGATGGCAATGCAATGCC 646
QY      440 oArgCysValGluSerThrAlaTyrcysGlyProProProSerIleAsnAsnGlyAspTh 460

```

```

Db      647 ACTGTCGTGCAAAATTCACA--TGTTACACACCACTAGATGAAGACGGAGCACT 703
QY      460 rThSer 462
      704 TAAATCA 710

RESULT 9
BG564962
LOCUS   602583332p1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711903 5',
DEFINITION
ACCESSION   BG564962
VERSION     BG564962.1 GI:13572615
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 912)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CM1552 row: o column: 08
            High quality sequence stop: 713.

FEATURES
     source          1..912
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4711903"
                     /clone_lib="NIH_MGC_76"
                     /lab_host="DH10B (TI phage-resistant)"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgctcgcc); Site_2: SfiI (ggcattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGCGCATATGAGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGCGCGAGCGCGCATG-3' (30)BP-3' (where B = A,
                     C, G, or T). Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH MGC Library."

BASE COUNT      296 a      187 c      173 g      256 t

ORIGIN
Alignment Scores:
Pred. No.:      1,63e-74      Length:      912
Score:          788.50      Matches:      176
Percent Similarity: 35.86%      Conservative: 18
Best Local Similarity: 32.53%      Mismatches: 43
Query Match:    24.77%      Indels:      304
DB:            12      Gaps:      2

US-09-805-337A-2 (1-569) x BG564962 (1-912)
QY      1 MetLeuLeuPheSerValIleLeuIleSerTTPValSerThyValGlyGlyGly 20
      84 ATGTGGCTCTCTGTCAGTGAATPCTAACTCAGGATATCCCTGTTGGGGGAGAGA 143
QY      21 ThrLeuCyAspPheProLysTleHisIsglyPheLeuTyraSpGluLysPtyrAsn 40
      144 ATGTTCTGTGATTTCCAAAATAAACCTGGAATTTCTATATGATGATAAATAATTAA 203
QY      41 ProPheSerGlnValProThrGlyGluValAlaPheTyTyrSerCysGlyTyrAsnPheVal 60

```

```

Db      204 CCATTTCGCAAGTTCCTACAGGAGAGCTTTCTATTACTGCTGTGAATATATTG 263
QY      61 SerProSerLysSerPheThrThrArgIleThrCysThrGluGlyTyrSerProThr 80
      264 TCTCCCTTCAAAATCCCTTTGGACTGCATAACTCCGAGAGAGATGTCACCAACA 323
QY      81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlySer 100
      324 CCAAGCTGTCTACAGCTGTGTCTTCTTTGTGGAAAATGTCATTTCTAACTTCA 383
QY      101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu 120
      384 GGACAAACACATCTGGAAGGTGATCTGTGACAAATATTATTCACACAGGATACAGACT 443
QY      121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTyrPheThrProProIleCys 140
      444 CAAACAAATGAAACACATTTCAATGCTGACAGCGGCTGTGTCCACGCTCCCAATGC 503
QY      141 SerPheThrLysGlyGlyCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160
      503 -----
QY      161 LysLysGlnSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180
      503 -----
QY      181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTyrPheProAsnProThr 200
      503 -----
QY      201 CysLysGlyGluValArgSerCysGlyProProProGlnLeuSerAsnGlyGluValLys 220
      503 -----
QY      221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240
      503 -----
QY      241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTyrThrLeuPro 260
      503 -----
QY      261 ThrCysValGluGluValLysThrCysGlyTyrIleProGlnLeuGluTyrGlyTyrVal 280
      503 -----
QY      281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAspCysArgAsnGlu 300
      503 -----
QY      301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyTyrIleThrGlnLeuPro 320
      503 -----
QY      321 MetCysValAlaIleThrHisGlnLeuLysArgCysLysIleAlaGlyValAsnIleLysThr 340
      503 -----
QY      341 LeuLeuLysLeuSerGlyLysGluPheAsnHisAsnSerArgIleArgTyrArgCysSer 360
      503 -----
QY      361 AspIlePheArgTyrArgHisSerValCysIleAsnGlyLysTyrAsnProGluValAsp 380
      503 -----
QY      381 CysThrGluLysArgGluGlnPheCysProProProGlnIleProAsnAlaGlnAsn 400
      503 -----
QY      401 MetThrThrThrValAsnTyrGlnAspGlyGlyLysValAlaValLeuCysLysGluAsn 420

```

Db 503 ----- 503  
 QY 421 TyrLeuLeuProGluAlaLysGluLeuValCysLysAspGlyArgTrpInsLeuPro 440  
 Db 503 ----- 503  
 QY 441 ArgCysValGluSerThrAlaTyrCysGlyProProPheSerIleAsnAsnGlyAspThr 460  
 Db 504 AGTCACACTATTCTGACAGAAAATGTGGCCCCCACCACCATGATGACATGACACAT 563  
 QY 461 ThrSerPheProLeuSerValTyrProPheGlySerThrValThrTyrArgCysGlnSer 480  
 Db 564 ACTTCACTTCTGTGTGACATATATGCTCCAGGTTATCATGAGTACAGTCCAGAAC 623  
 QY 481 PheTyrLysLeuGlnGlySerValThrValThrCys-ArgAsnLysGlnTrpSerGluPro 500  
 Db 624 TGTATCAACTGTGAGGTACAAATCAATATACATGTATAGAAAGCAGCATGTGACAAAC 683  
 QY 500 oProArgCysLeuAspProCysValValSerGluGluAsn---MetAsnLysAsnAsnIle 519  
 Db 684 ACCAAATGCTTAGATCCATGCTGTATATACACAGAAATTTATGACAAAATTTACCAT 743  
 QY 519 eGlnLeu--LysTrpArgAsnAspGlyLysLeuTyrAlaLysThrGlyAspAlaVal 537  
 Db 744 TAAATTTAAAGTGGACAAACCAACAAAGCTTTATTCAAGCACCGGTGACATAGTT 800  
 RESULT 10  
 LOCUS BG533258 786 bp mRNA linear EST 03-APR-2001  
 DEFINITION 601860633f2 NIH\_MGC\_76 Homo sapiens cDNA clone IMAGE:4072108 5',  
 mRNA sequence.  
 ACCESSION BG533258  
 VERSION BG533258.1 GI:13524798  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE NIH-MGC http://mgi.mc.man.ac.uk/.  
 AUTHORS 1 (bases 1 to 786)  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM918 row: m column: 05  
 High quality sequence stop: 762.  
 Location/Qualifiers  
 1..786  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4072108"  
 /clone\_lib="NIH\_MGC\_76"  
 /lab\_host="PHIOB (T1 phase-resistant)"  
 /note="Organ: liver; Vector: pDMR-LIB (Clontech); Site: 1;  
 5'flit (ggcgccgcgcgcgc); Site: 2; 5'flit (ggcgccatagcgc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCCATTAATGAGC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCGCGCCGACATG-dT(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.85  
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH-MGC Library."  
 BASE COUNT 264 a 157 c 146 g 219 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2,76e-74 Length: 786  
 Score: 785.50 Matches: 172  
 Percent Similarity: 36.654 Conservative: 23  
 Best Local Similarity: 32.338 Mismatches: 36  
 Query Match: 24.688 Indels: 302  
 DB: 12 Gaps: 2  
 US-09-805-337A-2 (1-569) x BG533258 (1-786)  
 QY 1 MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGly 20  
 Db 84 ATGTGGCTCTGTCGACGTATTAATCTTAATCCAGGATATCTGTTGGGGACAAACCA 143  
 QY 21 ThrLeuCysAspPheProLysIleHisGlyPheLeuTyrAspGluAspTyrAsn 40  
 Db 144 ATGTTCTGTGATTTCCAAAATTAACCTGGAATTCATATGATGACAAACAAATATAG 203  
 QY 41 ProPheSerGluValProThrGlyGluValPheTyrTyrSerCysGluTyrAsnPheVal 60  
 Db 204 CCATTTCCCAAGTCTCCACAGGAGAGTTTCTATTACTCCTGTCGAATATATTTTGTG 263  
 QY 61 SerProSerLysSerPheThrPheThrArgTyrThrCysThrGluGlyTyrSerProThr 80  
 Db 264 TCTCTTCAAAATCTCTTTGACCTGCATACGTGCGGAGAGGAGATGTCACCAACA 323  
 QY 81 ProLysCysLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGluSer 100  
 Db 324 CCAAGCTCTCCAGACTGCTGTTCTTCTTTGCGAAAGCGCATTTCTGATCTTCA 383  
 QY 101 GlyLeuIleHisLeuGluGlyAspThrValGlnIleIleCysAsnThrGlyTyrSerLeu 120  
 Db 384 GGACAAACACATCTGAGAGGTGATACGTGACAAATTAATTTGCAACAGCATACAGACT 443  
 QY 121 GlnAsnAsnGluLysAsnIleSerCysValGluArgGlyTyrSerThrProProIleCys 140  
 Db 444 CAAACCATGACAAACATTTTCATGATGTAGAACGGGCTGCTCCACTCTCCAAATGC 503  
 QY 141 SerPheThrLysGlyGlyCysHisValProIleLeuGluAlaAsnValAspAlaGlnPro 160  
 Db 503 ----- 503  
 QY 161 LysLysGluSerTyrLysValGlyAspValLeuLysPheSerCysArgLysAsnLeuIle 180  
 Db 503 ----- 503  
 QY 181 ArgValGlySerAspSerValGlnCysTyrGlnPheGlyTrpSerProAsnPheProThr 200  
 Db 503 ----- 503  
 QY 201 CysLysGlyGlnValArgSerCysGlyProProGlnLeuSerAsnGlyGluValLys 220  
 Db 503 ----- 503  
 QY 221 GluIleArgLysGluGluTyrGlyHisAsnGluValGluTyrAspCysAsnProAsn 240  
 Db 503 ----- 503  
 QY 241 PheIleIleAsnGlyProLysLysIleGlnCysValAspGlyGluTrpThrThrLeuPro 260  
 Db 503 ----- 503  
 QY 261 ThrCysValGluGlnValLysThrCysGlyTyrIleProGluLeuGluTyrGlyVal 280  
 Db 503 ----- 503  
 QY 281 GlnProSerValProProTyrGlnHisGlyValSerValGluValAsnCysArgAsnGlu 300  
 Db 503 ----- 503  
 QY 301 TyrAlaMetIleGlyAsnAsnMetIleThrCysIleAsnGlyIleTrpThrLeuLeuPro 320  
 Db 503 ----- 503







Db	425	AAATCAATCAATTTAAATTATATCTTGACGACACATTAAAAACAAGAGATTCGATTAT	484
QY	352	AsnSerArgIleArgTyrArgCysSerAspIlePheArgTyrArgHisSerValCysIle	371
Db	485	AATCTCAACAATAGAGTACAGATGTAGAGGAAGAAAGAGATGATATACACAGCTGCATTA	544
QY	372	AsnGlyLysTrpAsnProGluValAspCysThrGluLysAspGluGlnPheCysProPro	391
Db	545	AATGCAAGATGGATCTCCAGAGTCAATCGTCAATGGCACAATACATTAATGCTCCT	604
QY	392	ProProGlnIleProAsnAlaGlnAsnMetThrThrThrValAsnTyrGlnAspGlyGlu	411
Db	605	CCACCTCAGATTCCCAATTTCCACAATATGTACAAACCAACATCGAATTATTCGGATGGAGAA	664
QY	412	LysValAlaValIleCysLysGluAsnTyrIleuProGluAlaLysGlu	428
Db	665	CACAGTATCTGTCTTGGCAGGAATATTAATTCAGAGAAGAGAGAA	715
RESULT 13			
LOCUS	AV682232	815 bp	mRNA linear EST 16-JAN-2002
DEFINITION	AV682232 GKB Homo sapiens cDNA clone GKBAG04 5', mRNA sequence.		
ACCESSION	AV682232		
VERSION	AV682232.1	GI:10284095	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 815)		
	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.		
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)		
MEDLINE	21625106		
COMMENT	Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@hgc.sh.cn This clone is available at CHGC in Shanghai.		
FEATURES			
source	Location/Qualifiers		
	1..815		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="GKBAG04"		
	/clone_1lb="GKB"		
	/tissue.type="hepatocellular carcinoma"		
	/dev.stage="Adult"		
	/lab_host="SOLR"		
	/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	252 a 166 c 173 g 223 t	1	others
ORIGIN			
Alignment Scores:			
Pred. No.:	2 426-73	Length:	815
Score:	777.00	Matches:	152
Percent Similarity:	66.27%	Conservative:	17
Best Local Similarity:	59.61%	Mismatches:	79
Query Match:	24.41%	Indels:	7
DB:	10	Gaps:	3
US-09-805-337A-2 (1-569) x AV682232 (1-815)			
QY	1	MetLeuLeuLeuPheSerValIleLeuIleSerTrpValSerThrValGlyGlyGluGly	20

Db		62	ATGAGGCTCCGGTGCAGTGATATTCTATATCTCACCGCATATCCTGTGGGGGAGAACA	121
OY		21	Thrlucycasppheprolysilehisisclypheleutyraapgiugluasptryasn	40
Db		122	ACATTTGTGATTTTCACAAATAAACCATGGAATTCATATGATGAGAAGAAAATATTAAG	181
OY		41	prophesergrinalprothrnglyglualvalphenyltyrsercysgultyrasnpheval	60
Db		182	CCATTTTCCCAGGTTCTACAGSGGGAAGTTTTCATTTACTTCCTGTGAATATAATTTTTGTC	241
OY		61	SerProSerLysSerPheTrpThrArgIleThrCysThrnglugluclytTPserProThr	80
Db		242	TCTCCTTCAAAATCATTTTGGACGTCGCCATACATACAGACAGAGAAGATGGTCCACACA	301
OY		81	ProlYscYsLeuArgMetCysSerPheProPheValLysAsnGlyHisSerGlusSer	100
Db		302	CCAAAGTGTCTCACACAGCTGTGTTCTTCTTCCTTTGTGCAGAAATGCTCATTCGAATCTTCA	361
OY		101	GlyLeuIleHisLeuGluLYASPThrValisInllelecysasnThrGlyTySerLeu	120
Db		362	GGACAAACACATCTCGAAGGATGATCTGTCAAATTTATTTCCACACAGGATACAGACTT	421
OY		121	GLASnasncgluLysasnIleSerCysValgluaarglytPserThrProProIleCys	140
Db		422	CAAAACATGAGACACACATTTCTATGTGTAGAACGGGGCTGTCCACCCTCCCAATATGC	481
OY		141	SerPheThrLysglyglucyHisValProIleLeuGluAlaan-----Valasp	157
Db		482	AGTGCACATGACACTTCCTGTGTGAATCCGGCCACACATACAAATGCTTATATAGTGTGC	541
OY		158	AlaGlnProLysLysGluSerTyrLysValisValLysaspValLeuLysPheSerCysArgLys	177
Db		542	AGACAGATGAGTATAA-----TATCCATCTGCTGACAGACATGATCATCATGTAGACGC	595
OY		178	AsnLeuIleatrgvalglySerAspSerValalncystyrylglnPheglytTPserProbsn	197
Db		596	CCTTATGAATAGTTTGGGGATGAGAAAGTATGCTGTTAAATGGAATGGACGACCAACA	655
OY		198	PheProThrCysLysGlyGlnValArgserCysglYProproPoglnLeuSerFasnly	217
Db		656	---CCTCATGTCAAAAGATTCTACGGGAAATGTGGCCCCCTTCNACCATTGACAAATGGG	712
OY		218	gluValLysgluIleatrgLysgluLutrglyL-HisasnIluValglutryaspCy	237
Db		713	GACATTACTTATCTCCCGGTCTCAGGAATATCTCAGCTTATCAAGTGTGATGCCATGG	772
OY		237	sasnProbsnpheIleasnnglyprolylsylellelncys	251
Db		773	CAAGACTGTATCACTTGGAAGGACACACAGATACCATGT	815
RESULT 14				
LOCUS	AV688058	761 bp	mRNA	Linear EST 16-JAN-2002
DEFINITION	AV688058 GKC Homo sapiens cDNA clone GKDDM03 5', mRNA sequence.			
ACCESSION	AV688058			
VERSION	AV688058.1	GI:10289921		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,			
	Shao,H., Qu,D., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,			
	Xiao,K., Yu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,			
	Hu,G., Gu,J., Chen,Z., and Han,Z.			
TITLE	Insight into hepatocellular carcinogenesis at transcriptome level			
JOURNAL	with those of corresponding noncancerous liver			
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)			
COMMENT	21625106			
	Contact: Zenguang Han			





OY 500 oProAtgCysLeuAspProCysValValSerGIuGIuAsnMetAsnLysAsnAsnIleGI 520  
 Db 669 AGCAAAATGCTTAGATCATGCTATATATCAAG--AATTATGGAATAATATACATPAA 726  
 OY 520 nLeuLys 522  
 Db 727 ATTTAAG 733

Search completed: July 23, 2003, 09:39:43  
 Job time : 2209 secs

**THIS PAGE BLANK (USPTO)**